

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 13:56:39 ; Search time 6938.09 Seconds
(without alignments)
8578.019 Million cell updates/sec

Title: US-09-776-865-3
Perfect score: 2844
Sequence: 1 cccggggggcggggggtctc.....ataaaaaaaaaaaaaa 2844

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sy: *
12: gb_un: *
13: gb_vl: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgc_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	2844	100.0	2844	4	AF244578	AF244578 Ovis arie
2	2844	100.0	2844	6	AX207626	AX207626 Sequence
3	1572.4	55.3	2512	6	AX138494	AX138494 Sequence
4	1572.4	55.3	2512	9	HSAB37747	AJ387747 Homo sapi
5	1570.8	55.2	2930	6	AX207624	AX207624 Sequence
6	1570.8	55.2	2930	6	AF244577	AF244577 Homo sapi
7	1570.8	55.2	3329	9	BC020961	BC020961 Homo sapi
8	792.2	27.9	1587	9	AK025880	AK025880 Homo sapi
9	602.8	21.2	2006	9	AK026921	AK026921 Homo sapi
10	480.2	16.9	56641	2	AL138833	AL138833 Homo sapi
11	480.2	16.9	113202	9	HS397H23	AL121972 Human DNA
12	480.2	16.9	149597	2	AC034471	AC034471 Homo sapi
13	480.2	16.9	157749	2	AC025535	AC025535 Homo sapi
14	314.2	11.0	375	6	AX118967	AX118967 Sequence
15	269.6	9.5	3616	9	AK024903	AK024903 Homo sapi
16	268	9.4	2626	2	AB020527	AB020527 Homo sapi
17	261.4	9.2	179129	2	AC097023	AC097023 Rattus no
18	230.2	8.1	2004	10	BC013445	BC013445 Mus muscu
19	228.6	8.0	1885	10	MMNP1CT	X77241 M.musculus
20	226.6	8.0	1841	3	AY060776	AY060776 Drosophill
21	225.4	7.9	3946	9	AB032435	AB032435 Homo sapi
22	223.2	7.8	1700	10	RNU28504	U28504 Rattus norv
23	219	7.7	1865	10	BC018252	BC018252 Mus muscu
24	218	7.7	2528	10	AF324864	AF324864 Mus muscu
25	217.4	7.6	3982	10	AF271235	AF271235 Rattus no
26	210.2	7.4	1549	6	AX336195	AX336195 Sequence
27	210.2	7.4	1549	6	AX336653	AX336653 Sequence
28	210.2	7.4	1549	9	HSNAP11	X71355 H.sapiens m
29	209.8	7.4	1874	10	BC018306	BC018306 Mus muscu
30	205.4	7.2	1794	9	HUMAP14	D28532 Human mRNA
31	193.2	6.8	2024	10	RNU07609	U07609 Rattus norv
32	192.6	6.8	2748	3	AY069501	AY069501 Drosophill
33	190.6	6.7	1710	3	AY075277	AY075277 Drosophill
34	190.6	6.7	34229	2	AC014600	AC014600 Drosophill
35	190.6	6.7	267488	3	AE003640	AE003640 Drosophill
36	190.6	6.7	303885	9	DROSADH01	AE003407 Drosophill
37	186.8	6.6	2366	6	AB032436	AB032436 Homo sapi
38	185.2	6.5	2716	6	I40028	I40028 Sequence 1
39	185.2	6.5	2716	6	I40029	I40029 Sequence 3
40	185.2	6.5	2716	6	I40220	I40220 Sequence 1
41	185.2	6.5	2716	6	I40221	I40221 Sequence 3
42	185.2	6.5	2716	6	I73259	I73259 Sequence 1
43	185.2	6.5	2716	6	I73260	I73260 Sequence 3
44	183.6	6.5	2178	3	AF095787	AF095787 Caenorhab
45	180.4	6.3	1855	4	RABRCTX	M76466 Rabbid rena

ALIGNMENTS

RESULT 1
AF244578
LOCUS AF244578 2844 bp mRNA linear MAM 06-AUG-2000
DEFINITION Ovis aries membrane glycoprotein SP55 (sp55) mRNA, complete cds.
ACCESSION AF244578
VERSION AF244578.1 GI:9719375
KEYWORDS
SOURCE
ORGANISM
Ovis aries
sheep.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Ovis.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yao,H.-P.,
Carter,C.E., Shl,E., Venkov,C., Yakes,M.F., Page,D.L. and H.C.G.
Identification of a novel membrane protein from mammalian cells
that interact with the anti-pathoangiogetic compound CM101
unpublished
2 (bases 1 to 2844)
Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R. and Hellerqvist,C.G.
Direct Submission
Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School
of Medicine, 23rd@Pierce, Nashville, TN 37232-0146, USA

```
FEATURES
source
Location/Qualifiers
1..2844
/organism="Ovis aries"
/db_xref="taxon:9940"
/cell_type="endothelial cell"
/tissue_type="lung"
1..2844
/gene="sp55"
84..1571
/gene="sp55"
/feature="similar to Na+ inorganic phosphate transporter"
/codon_start=1
/product="membrane glycoprotein SP55"
/protein_id="AAB97770.1"
/db_xref="GI:9719376"
/translacion="MKSPPSDDLAPSDGEESDRPTLLIQRAPRAPEAPVCCSARYNIAP
LSPEGFVLXSLAPVNLVSLAVDMDNTRAKDRRTYSECEHSAPIKVLHNOTGKRY
WDAETOGWILGFFFGITITQIPGIVASHSGKLLGFEIFATLFTPLAADFG
VGAIVLRALIELEGVTPAMHAMSVPLEKSLISIYAGOLGVSLPLSG
VICYVNMVYVEYFEGIVGILMFILMICVSDPEPKTTPYEKEIISLKNLSS
OKSVPMIPMLKSLPLMAIVVAHFSYNMFTLLTLPYKELVRENIENGFLSAP
YLGCWLCMLISGOADNLRARMSSTLWRRVSLIGMIPALFELVAAGTIGDYSIA
VARLTSTLIGRCSGSFSINHDIAVSAGILGITNTPATIPGMIGPIIASLPIE
NTIGENOTVPCIAAINVFAIFFTTLPAGEVONMAISDHQHRN"
BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN
Query Match 100.0% Score 2844: DB 4: Length 2844:
Best Local Similarity 100.0% Pred. No.0:
Matches 2844: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1 cccggggcggggggcttcggcggtccgcgtgagcctctcttcgcggagagcagtttg 60
Db 1 CCCGGGGCGGGGGGCTTCGGCGGCTCCGCGTGGAGCTCTTTTCCGGAGCAGTTTG 60
QY 61 cgcgcgtagctccctgaagcattgaagtcgccggtttcgagcttagcccgagagc 120
Db 61 CGCGGTAGCTCCCGAAGGAGTCATGAAGTCCCGGTTTGGACTTGAAGCCCGAGCAGC 120
QY 121 gcgagggagggctggagcagcaccgcgtctctgcagcgcgcccgcgaggaacgcgtc 180
Db 121 GCGAGGAGGGCTGGAGCAGCACCCTCTCTGCAGCGCGCCCGCGGAGGAGACCGCTC 180
QY 181 cagtaatgcgtcgtcgttacaacacctagacatttctcttcttcttggttcttcg 240
Db 181 CAGTATGCTGCTGCTGCTGTACAACTTGTCTCTTTTGGTTTCTTCTCTTC 240
QY 241 tctattcaatcaggggtgaatcctgagcgttgcaatgtgagcattgtgatccaacaa 300
Db 241 TCTATTCAATCAGGGTGAATCTGAGCTTGCACATGAGTGAATGATGATTCAAACAA 300
QY 301 ctggcaagaataatagaagctccctaagagtgctgagagcattctgctccataaagttc 360
Db 301 CTGCGCAAGATTAATAGACGCTCTACAGATGTGCAAGACATTCGCTCCCATTAAGATTC 360
QY 361 ttcacaacccaacgggttaaaagtaaccggtgagatgagagaacactcaagagatgcg 420
Db 361 TTTACAAACCAACGGGTAAAGTACCGGTGGATGCAGAAACTCAAGAGATGATTCG 420
QY 421 gatcttttctatgctacatcatcaacaattcctgagagatattgttgcagcagaa 480
Db 421 GATCTTTTCTATGCTACATCATCACAAATTCCTGGAGGATATGTTCAGCAGAA 480
QY 481 gtggggggaagcgtgtgctggatcgggattcttgctaaagctattcaacctgttca 540
Db 481 GTGGGGGGAAGCTGTGCTAGGATTCGGGATCTTGTCTACAGCTATCTTCAACCTGTTC 540
QY 541 ctccccctgcgtcagattcggagcttgcagccctgttgcactcaagggcactagaagggc 600
Db 541 CTCGCCCTGCCTGAGATTTTGGAGTGGAGCCCTGTGTGCACTCAGGCGACTAGAAGGCG 600
QY 601 taggagaggggtgcacataaccagcaatgcatcgttctcattcattggtccctcc 660
```

```
|||||
Db 601 TAGAGAGGGGTGCATATCCAGCCATGCATGCATGTGGTTCATGTGGCTCCCTC 660
QY 661 ttgaaagaagcagcttcgtgatttccatagtgaagagcaaaccttggagcagtaattc 720
Db 661 TTGAAAGAAGCAGCTTCGTGATTTTCATATGAGAGCAACAACCTTGGGACATGATTT 720
QY 721 cttctctcttcttgagtaatttgcataatagaattggaactatgcttctattct 780
Db 721 CTCCTCTCTTCTTGAGATTAATTTGCATATAGAAATTTGACTATGTCTTCTATTTCT 780
QY 781 ttggcattgttggaatcaatcgtgttattttagatgcgttagttagtataaccag 840
Db 781 TTGGCATTTGTGAATCATCTGCTTTATTTATGACTCTCTTATGATGATACACAG 840
QY 841 aaactcaagaacatactaccctccatbaaagaagtatcttcttcaatcaaaaaatc 900
Db 841 AAATCTCAAGACATATACCTCCATGAAAAAGAGTATATCTTTCATATTAATAATC 900
QY 901 agctctctcaagaagtcagtgccgtggaatacctatgcgaatcaactgcacattgg 960
Db 901 AGCTCTCTTCAAGAAATCATGTGCCGTGATACCTATGCTGAATTCACCTGCACTTGGG 960
QY 961 ctattgtcgttgcacatttctttaaaccgtgacatttatacttctgtgaacctatgc 1020
Db 961 CTATTGTGCTTGACATTTTCTTACAACTGAGACTTTTATCTTGTGACCTTATTTGC 1020
QY 1021 ctacttcaatgaagaagtcctaaagttcaatctcaagaagaatggcttttcttcgag 1080
Db 1021 CTACTTCAATGAGAGAGTCCCTAAGTTCAATTTCAAGGAAGAGGTTTATCTGACG 1080
QY 1081 tcccttattagtttgttctgttctatgtaagcttctgtcgtcgaagctgcgaacattaa 1140
Db 1081 TCCCTTATTAGGTTGTGTGTTATGATGATCCGTGCGGTCAAGCTGTCAGCAATTAA 1140
QY 1141 gggcaagaatgaatttctcaactctgtgggttcgaagaattttagccttaaggagatga 1200
Db 1141 GGGCAAGATGAATTTTCAACTCTGCGGTTGGAACAGTTTATAGCTTATAGGGATGA 1200
QY 1201 ttggaacctgcgatatctctgttgcgagagatttaagcttgcatttctcttcgctg 1260
Db 1201 TTGGACCTGCGATATTCCTGTGTTGCCAGAGATTTATAGCTGTGATTTCTCTGGCTG 1260
QY 1261 ttgcattcctaaccataccaacaaccttggagagcttctcttcttcttgaattagaatca 1320
Db 1261 TTGCATTCCTAACCATATCAACAACCTGGAGGCTTTGCTCTTGATTAAGATCA 1320
QY 1321 accactcggacattgctcctctgtatgcgtgtattctccttgcggacatcaaatcctttg 1380
Db 1321 ACCATCTGGACATTTGCTCTGTATGCTGTATTCCTCGGGCATCCAAATACCTTTG 1380
QY 1381 ccaactatctctgaatgatgtggggcccaatcttccaagaagcttcaaccctggagaacatga 1440
Db 1381 CCACATTCCTGGAATGATTTGGGCCCATCATTTGCCAGAAATCTTACCTCGAGAACCTGA 1440
QY 1441 ttggaagaatggcaaacgttttctgcatacgtctgcatacaatgtaatttgggtgcattc 1500
Db 1441 TTGGAGAAATGGCAAACTGTTTCTGCATCCGCTGCTATCAATGATTTGGGCCATT 1500
QY 1501 tcttcaacatactgcgcaaaagtgaagtgcaaaactggtggccatcagtgatcaccagagac 1560
Db 1501 TCTTCAACATACTTCGCAAAAGTGAAGTGCAAAACGTGGCCATCAGATCACCAAGAGAC 1560
QY 1561 acagaacatgaagaacccaataataatcctgtcctcattaaatgtaattcttgttttcaat 1620
Db 1561 ACAGAACCTGAAGAACCAATTAATATCTGTCTATTAAATGATCTTGTATTCAT 1620
QY 1621 gtaacctaaagtcctcttataatttaagtgttaagcaatccataataagaataaat 1680
Db 1621 GTAACCTAAAGTCCCTTATATATTTAATGTTAAGCAATCTATATACAAAGTAAATTT 1680
QY 1681 gtaatagaataatgtgttagatttgaagccttgtaataatgaatgttcaactagtgc 1740
|||||
```

Db 1681 GTACTAGAAAAATTTGTTAGATTGTTGAAGCTGTGTAATCATGAAATGTCACACTGTC 1740
Oy 1741 atataagaacaaatagctatttttaattatattatataaccggttgcgtgaacttaattc 1800
Db 1741 ATATAGCAAAATATAGCTATTTTATATATATATTAACCGTTTGGTGGACCTTACAATTC 1800
Oy 1801 agggtcacatactggtctgaagtcgaagccaacccaatlaggggagttctattat 1860
Db 1801 AGGGTCACATATCTGGCTGCAAGTCAGGCAACCCACAATAGGGGAGTCTATTTATTTAT 1860
Oy 1861 aagaccataccctaaagaatagctgtaaatagacccttctaacttgcgttaataag 1920
Db 1861 AAGACCATACCTAAAGAAATGAGCTGAAATAGACCCCTCTATACCTTTGCTTAATTAAG 1920
Oy 1921 tgaataaatctcaagctctgttaaacatctgtttctgaaccctcccaaaaatt 1980
Db 1921 TGGATAAATATCTCAGGCTTTGTTAAACATCTGTTTGTGACCTTCCTCAAAAAAT 1980
Oy 1981 attgtcacaagcaatccctgaacatgtagtctcaacttgaacctcccaagagctg 2040
Db 1981 ATTGTCAATGACATCCCTGACATGATGCTCAAACTTTAGCTCTGACAGAGCTGG 2040
Oy 2041 cagccactgatactcaagctggaacttcaactgaggggaagcaatgccagagctgc 2100
Db 2041 CAGCCACTGATACATTCAGCCTGGCACTTCACTGAGGGAAACATGCCAGCAGCTGCC 2100
Oy 2101 acaatgccctctctggtctcaaggagctgcccagcaacttagagagatccaaagcag 2160
Db 2101 ACATGTCCCTCTCTGCGCTTCAAGGACAGTGCACCACTTAGGACAGATCCAAAGCAG 2160
Oy 2161 ggtcagcgccaagccttgaacggtatcttcccttggggcgtgtaactgctgtaag 2220
Db 2161 GGTCAAGCCCAAGGCTTTGAGAGGTATCTTCCCTGGGGGCTGTAATGTGATGAAG 2220
Oy 2221 cccgtagcacaagggagcaagcgcatccacagatcgtttccatgcaacctctccctc 2280
Db 2221 CCGTAGGCAACAGGGAGCGCGCATCCACATCATGTTGTTCCATGACACCCCTCCCTTC 2280
Oy 2281 ccttcacagcaactgtagatctgctgcatgtaactcgtcaaaaagaagtgtgagct 2340
Db 2281 CCTTCCAGACACTGGAGTATGGCTGGCATGTAACTGCAAAAGAAAGTGTATGCT 2340
Oy 2341 aattagcacacataacatcatccttgaatgacatcctcaacttcaatgatacagatataa 2400
Db 2341 AATAGCCACATATATACATCATCTTGATGATCCTTACCTTCACATGATGATGATATAA 2400
Oy 2401 tcttaaatctctggtcttagagagctacacagaataataataataaagaacaaa 2460
Db 2401 TCTTCAAGTCTGTTGTAGAGCTTACACCAAGATAATTAATAAATAAAGAAACAAA 2460
Oy 2461 acaattttctgctgacacctaagtgctggtgacgttcaaggtttaaagtgaactcta 2520
Db 2461 ACATTTTCTGCTGACACCTTAAGTGTCTGTTCAAGTTCAAGTTAAAGTGAATTTCTA 2520
Oy 2521 ctcaacataacctgcaacccggtgtaacatcctttagtgttcttcaaatcttat 2580
Db 2521 CTTCACATACCTGCAACCGGTGTATATCATCTTATAGTGTGTTCTTAAATCTTAT 2580
Oy 2581 ttttcaattttctcgtgacacatctcgaagtggttttagacatgltttaggggacttat 2640
Db 2581 TTTTCAATTTTCTCTGACCATCTTCCAGTGTGTTTGAAGTGTGTTGAGGGCAATTTAT 2640
Oy 2641 gtgatttgaacctgatttaagtgttcaactggtgtaacttcaaacactgtaataattta 2700
Db 2641 GTGATTTGAAGCTGATTAATGTTTCACTGTGTATGTTCAACACTGATTAATTTTAA 2700
Oy 2701 actaaagctatttaactgataatgagtgtatatactctgtaaatlttttaaatctg 2760
Db 2701 ACTAAAGCTATTATTAATGAATATGATGTATATCATCTGTAATTAATTTTAAATCTG 2760
Oy 2761 taaatagtttaagtgtgataagtgataatttcttllacaataacaataaatcttcttg 2820
Db 2761 TAAATAGCTTTTAAGTTGCTATGTGTATTTCTTTTACAAATCAAAATTAATCTTTTGG 2820

Oy 2821 aatgataaaaaaaaaaaaaaa 2844
Db 2821 AATGATAAAAAaaaaaaaaaa 2844
RESULT 2
AX207626
LOCUS AX207626 2844 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 3 from Patent WO0156598.
ACCESSION AX207626
VERSION AX207626.1 GI:15422331
KEYWORDS
SOURCE Ovis sp.
ORGANISM Ovis sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprine; Ovis.
REFERENCE 1 (bases 1 to 2844)
AUTHORS Hellerqvist, C.G.
TITLE Methods for preventing or attenuating pathoangiogenic conditions by
using the gbs-toxin (cm101) receptor as a vaccine
JOURNAL Patent: WO 0156598-A 3 09-AUG-2001;
VANDERBILT UNIVERSITY (US)
FEATURES
source location/Qualifiers
1: 2844
/organism="Ovis sp."
/db_xref="taxon:9939"
84..1571
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60359.1"
/db_xref="GI:15422332"
/translation="MKSPPSLAPSDGEGSDRPILOARAPAPAPVCCSARYNLAR
LSRFGFVLYSLRNLVSLALVDWDSNTAKDNRTSYCAEHSAPIKVLMOTGKKR
WDAETQWILDSFEGYITITDIPGGVYASRGGKILGFIATATFLPLADFG
VGAVALRALTEGLEGVTPAMHMAWSSWAPLESKILSYVGAOAGTVASPLSG
VICYMMNTVYFVFFGIIVGILIMFLIMICLVDPETHETIPYKEYILSLKNOSS
OKSVPMIDMLKSLPLMAIVVAHFSYNMFTVTLTLPTMYKEVLRFNIOENGPIASVP
YLGGICMLISGDAADNRARNFSYLVAVRVSILIGMIGAFILYNAAGFIGDYSILA
VAFITSTTLGFCSSGFSINHLIDAPSYACILIGITRTPTFMIGPIIARSLTPE
NTIGEMQVTFICIAAINVEGAIFETLFAKGEVQNWALSDHGHNR"
BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN
Query Match 100.0%; Score 2844; DB 6; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 cccggggcggggggcttcggcggtcccgctgagctctcttccggcgagcaggttg 60
Db 1 CCCGGGGCGGGGGCTTCGGCGGTCCCGCTGAGCTCTCTTTCCGCGAGCAGGTTTG 60
Oy 61 cgcgtagctcccttgaaggacatgaagttcccggttgcgacttagcccgagcgagc 120
Db 61 CCGCGTAGCTCCCTGAAGGACATGAGAGTCCCGGTTTCCGAGACTTAGCCCGAGGAGC 120
Oy 121 gcgagagaggtccggaacgcgacacgcgtctctgcaagcgccccggcgagacccgctc 180
Db 121 GCGAGAGAGGCGCGGACCGCACCGCTCTCCACAGCGCCCCGGGGGGGAACCCGCTC 180
Oy 181 cagtagtgcctcgtcgtgacacacacagcaatttctccttttggtttgccttcgttc 240
Db 181 CAGTAGTGTCTGCTCTGCTGTACCACTTAGCAATTTTGTGCTTTTGGTTTCTTCGTTTC 240
Oy 241 tctattcatiaaggtgaatctgagcggttgcactagtggaacatggttgatcaaacacaa 300
Db 241 TCTATTCAATTAGGGGTGAATCTGAGCGTGTGCACTAGTGGAACATGTGGATTCAAACACA 300
Oy 301 ctgcccaagataataagaacgtctcaagagtggtgcagagcatctcgtcccataaagctc 360
Db 301 CTGCCAAAGATTAATGAACGTCTCAAGTGTGAGAGCATCTCTCCCATAAAAAGTTTC 360

QY	361	ttcacaccaaacggygtaaaaagtaaccggtggyga tgcagaactcaaggatgaltctcg	420
Db	361	ttcacaaacaaacgggtataaaagttaccggcgagatccagaaactcaagatgatattctcg	420
QY	421	gactctttttctatggtcctaacatcaacaagtctctgagaatattgltgcagaaga	480
Db	421	gattctttttctatggtcctaacatcattcaacaaattcctgagatattggtgccacagaa	480
QY	481	gtggggggaagctgtcgtaggaatcgggaatcttgctaacagatcattcaacctgttca	540
Db	481	gtgggggggaacctgtctgctaggaattcgggagattcttgctaacagatcattcaacctgttca	540
QY	541	ctccctctgcgcgaagatttcggagtcggagacctgtgtgactcaggagactagaagggc	600
Db	541	ctccccctgcgcgaattttggagttcggaactcggaacctgtgtgactcaggagactagaagggc	600
QY	601	taggagagggtgtacataltccagccaatgacatgtgtgtcttcattgtgtccctc	660
Db	601	taggagagggggtcacattacattccagccatgcatgcatggtgttcatgtggctccctc	660
QY	661	ttgaaagaagcaagctcttcgaattatcatatgcagagagcaaaacttggagacagattt	720
Db	661	ttgaaagaagcaaaccttctgcagatttattatgtatgcagagagcaaaacttggagacagattt	720
QY	721	cccttcctcttcggaggtaatgtgcactaatgtgaattggagctatgctctattct	780
Db	721	ctcttccctcttcttggaggtaatgtgcactaatgtgaattggagctatgctcttattct	780
QY	781	ttgcatctgtttgaatcatcctgylttattatgtatcctgtagttagtgaataccag	840
Db	781	ttggcattgtttggagatcatctggtttattttatgagatcgtgttagttagtgaataccag	840
QY	841	aaaccacaagaacaatcaaccctccatbaaagaagatattcttcatcattaaaaact	900
Db	841	aaacgcacaagaacatcatctacctcctatgaanaagagatattcttcatcattaaaaatc	900
QY	901	agctctcttcaaaagtcagtcgccgtgtgatacctatgtcgtgaatcaactgcacctttgg	960
Db	901	agctctcttcaaaagtcagtcgccgtgtgatacctatgtcgtgaatcaactgcacctttgg	960
QY	961	ctatgttcgttgcaacatttttcttaaacctgcgactttatacttgttgacctattgc	1020
Db	961	ctatgttcgttgcaacatttttcttaaacctgcgactttttatctgttgacctattggc	1020
QY	1021	ctactacatbaaagaagctccatgaagtgtcaaatctcaagaagaatgggtttatctcag	1080
Db	1021	ctacttactatbaaagaagctccatgaagtgtcaaatctcaagaagaatgggtttattctcag	1080
QY	1081	tcocctattaggttgtgtgtatgtatgcacccgtgcgggtcaagctgcgtgcaaat	1140
Db	1081	tcocctattattaggttgtgtgtatgtatgcacccgtgcgggtcaagctgcgtgcaattttaa	1140
QY	1141	gggcaagaatgaatcttttaacactctgtgttcggaagaatcttttgaccttaagaagta	1200
Db	1141	gggcaagaatgaatcttttttaacactctgtgttcggaagaatcttttttgaccttaagaagta	1200
QY	1201	ttggaacctgcatacttcctgtgttcgcgcaggaattctaaagctgtatattcctctgctg	1260
Db	1201	ttggacacctgcgatatttctctgtgttcgcgcagagattttatagctgtgatatttcttgctg	1260
QY	1261	ttgcaatcttaacatatcaaaaccttggagactttgtcctctctcgtgaatttagcatca	1320
Db	1261	ttggcattctcttaaacattcaaaaccttggagagcttttggccttctctgattttagcatca	1320
QY	1321	aacacttggagaatgcctctctgtatgcgtgtatcttccttgggaatcaaaatattcttg	1380
Db	1321	aacatttggagaatgtcctctctgtatgcgtgtatcttctcttgggcatcacaataattcttg	1380
QY	1381	ccactatctctgaaatgattgggcccacatctgcagaaagtcttaacctgtagaacacta	1440
Db	1381	ccactattctctgaaatgattgggcccacatattgcccgaagatttttccctgtgaacacta	1440

QY	1441	ttggaagaatgccaacatgcttctctgcgaatgcgcgcgtctatcaatgtaatttggtcacatt	1500
Db	1441	ttgggaagaatggccaacactgttttctgcgacccctcctctatcattgattttggctccattt	1500
QY	1501	tcttcacactctctgcgaagaagtgtaaaagtgaacaacgaggccatcgaagtatcaccaagac	1560
Db	1501	tctttacaccttttggcccaaaagtgaaagtgcaaaactggggccatcagtgatcccaaganc	1560
QY	1561	acagaacatggaaggaacccaataaataccctgctctatataatgtaactcttgttatcat	1620
Db	1561	acagaaactggaaggaacccaataataatctctgctctattatgattgattttgttatcat	1620
QY	1621	gtaaccttaaaagtgcccttggatatttaatgtyttaaagcaactctataacagaataaact	1680
Db	1621	gttaacctaaagtgcccttggatatttttaattgaatgttaagcaatctaaatcagaatcttaatt	1680
QY	1681	gtacacagaanaaatgtgttgatgttgaagcgtctgaacatgaatctgaactgttgc	1740
Db	1681	gtactaganaanaattgtttgatgtttgttaagcgtctgaacatgaanaatgtcactgattgccc	1740
QY	1741	ataataagcaaaaattagctatttttaattatatacctaaccgcttctggaactataaactc	1800
Db	1741	anataagcaaaaatttagctatttttttaattattatttaaccgcttctggaacttaacaattc	1800
QY	1801	agggtcacatactctgctgcgaagtcaggcaaccacacaataaggsgagctctatattat	1860
Db	1801	agggtcacatattctgcgtgcgaagtcaggcaaccacacaataaggsgagctctattattatt	1860
QY	1861	aagaccatacctaagaagatgagctctgaataaagaaccttataactttgtaattaaag	1920
Db	1861	aagaccatacctaagaagatgagctctgaataaagaaccttataacctttgcttaattaaagg	1920
QY	1921	tgggaataaactctccagctctgtttaaacaatcgttttctgtaacactctcccaaaaaat	1980
Db	1921	tgggaataaataattcttcagctctgtttaaacaatcgttttctgtaacactctcccaaaaaatt	1980
QY	1981	attgtctacgaacatccctcgacatgtaagtlctcaaatcttagcctctccacgagctg	2040
Db	1981	attttgtctatcagcaatccctcgacatgtagtctcaaaacttttagcctctccacgagctgg	2040
QY	2041	cagcacagctatacttaagctctggaactgaactgaagggagaacatgcgcgaagcagctg	2100
Db	2041	cagcacacagctatcatttaagctctggaactgaactgaagggagaacatgcgcgaagcagctg	2100
QY	2101	acatgtcccccctctctgctcgaaggacagtgcccaacacttagcgaacatcccaagacag	2160
Db	2101	acatgtccccctctctgctcgaaggacagtgcccaacacttagtagcgaacatcccaagaccag	2160
QY	2161	ggtcagcgcgaagccttttgacggtatcttcccttggggcgttaatgtytggatgaag	2220
Db	2161	ggtcagcgcgcgaagccttttgacgsgtatcttcccttggggcgttaattgtggaatgaag	2220
QY	2221	ccctgaagccaacagggacagcgcgataccacaagtcaatgtttccatgtgacccctcccttc	2280
Db	2221	ccctgaagccaacagggacagcgcgataccacagatcagatgattttccatgacacccctcccttc	2280
QY	2281	cccttcacgaacacttggaatgattgctctgcatgttaaactgcgaanaagaagtgtgatcct	2340
Db	2281	ccctttccacgaacacttggaatgattgctctgcatgttaaactgcgaanaagaagtgtgatgcct	2340
QY	2341	aattagccaacataacatcatccttggatgactcactccttcacatgatacagatataaa	2400
Db	2341	aatttagccaacataataacatcatctcttgatgattgactccttcacatgatacagattataaa	2400
QY	2401	tcttcaagctcgtgtctctagaagctacaccagaaataattaaataaaaaagaacaaa	2460
Db	2401	tcttcaagctcctgtgtctctagaagctacaccagaaataattataataataaaaaagaacaaa	2460
QY	2461	acattttctcgtctctgacccaatgaagtgcttggttgatgattcaaggtttaaatgacttcta	2520
Db	2461	acatttttttctcgtctcagacccttaagtgcttggttgatgattcaaggtttaaatgacttcta	2520
QY	2521	cttcaataaccttcgaacccggtgtgtataatcattttagtgltgttttcttcaaatcttat	2580

Db	2521	CTTCACATACCTGGACCGGTGGTGATATCATCTTTATGTTGGTTCTTTAAACTTAT	2580
OY	2581	ttttccagttttcttcggaaccacttcgaagttggtttgacagctcttgaagcatttat	2640
Db	2581	TTTTCCAGTTTTCCGTGGACCATCTTCCAGTGGTTTGAAGCATGCTTTGAGGCAATTAT	2640
OY	2641	gtgattcgaactcgtattaatggttcaactcgtatgcttcaacactacctcgtaatattta	2700
Db	2641	GGTATTTTGGAACCTTATTAATGTTTACAGTGTATGTTTCAACACTACCGTGAATATTTTTA	2700
OY	2701	actaaagctatttaatgtaataatggtatagaacttcgttaataatattttaactcg	2760
Db	2701	ACTAAAGCTATTTAATGTAATATGATATGATATCAATCTGTAATTAATTTTAAATCG	2760
OY	2761	taaatagctttaagtctatgtaataatcttcttacaacatcaaatataactcttttg	2820
Db	2761	TAAATAGCTTTAAGTGTGATATGATATTTCTTTACAAATCAAAATTAATCTTTTGG	2820
OY	2821	aatgataaaaaaaaaaaaaaa 2844	
Db	2821	AATGATTAATAAAAAAAAAAAAAA 2844	

RESULT	3
AXI38494	
LOCUS	AXI38494
DEFINITION	Sequence 2 from Patent EP1069184.
ACCESSION	AXI38494
VERSION	AXI38494.1 GI:14274389
KEYWORDS	
SOURCE	human.

REFERENCE
1 (bases 1 to 2512)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE	JOURNAL
Human anion transporter gene implicated in salla disease and lysosomal stalic acid transport	Patent: EP 1069184-A 2 17-JAN-2001;

FEATURES	Location/Qualifiers
source	1. . 2512

```
BASE COUNT      612 a      585 c      566 g      749 t
ORIGIN          /organism="Homo sapiens"
                /db_xref="taxon:9606"
```

Query Match	55.3%	Score 1572.4	DB 6	Length 2512
Best Local Similarity	82.1%	Pred. No. 0		
Matches 1917, Conservative	0	Mismatches 386	Indels 33	Gaps 8

OY	5	gggggagggagggcttcgcgggctccggatctctcttccgcggagcaaggtttgcgc	64
Db	194	gccccgggggctttggggctgtcgagccggcgccctcttcttgcgaagtgagcagtracac	253
OY	65	gtagctcccttaagagcatcatgaagtcgccgggttcggacttaagcccgagcgacggcga	124
Db	254	CT--GCTACGTAGGCGGCATGAGAGTGTCCGCGTTGAGACCTGGCCCCGGAACGATGGCA	311
OY	125	ggaggggtcggagacggcaaacgcgtctctgcagcggcgcccgcgggcggaacccgctcaat	184
Db	312	GGAGGACAGGAGCCGACAGCCTTCTTTCACCGGGGGCCCCAGGGGCGAACCCTGCTCAGT	371
OY	185	atgcgtccttcgtcgtttacaacaccagcatcttttgccttttttgattcttcgtctcta	244
Db	372	GTGCGCTCTGCTGCTGTTACCACTTAGCAAAATTGTGGCCTTTTGTGTTCTTCATTGGTA	431
OY	245	ttcatataagggtgaatctcgaagcgtttgcactagtgtgacatcgttgaaatccaacaactgc	304
Db	432	TGCATTACGCTGATGATCGAATGTTGTGCTTAGTGGAATAGGTAAATTCCAATACACCTTT	491

QY	305	caagaataatagaaacgtcctcaagagtgctgcaagagcattctgtctccataaagaattctta	364
Db	492	AGAGATTAATGAACTTCCAAAGGGCTGTCCAGAGCATTTGCTCCCATTAAGATTCATCA	551
QY	365	caaccaaacgggtaaaaaagtaaccggtggagtgaagaacctaagagtgatctcgtgac	424
Db	552	TAATCAAAAGGGGTAAAGAAAGTACCATGAGATGAGAAACTCAAGAGATGAGATTCGCTTC	611
QY	425	ttttctctatgtctacatcaacaacaattctcctggagatgatgtgcaagcaagtg	484
Db	612	CTTTTTTATGTGCTACATCATCACACAGATTTCTCGGAGGATFATGTTGCCAGAAATATAG	671
QY	485	ggggaagctgtgtctagatctgggactctgtctacagctatctcaacctgtcactcc	544
Db	672	GGGGAAATGCTGCTAGAGATTTGGGATTCCTTGGCACAGTGTCTCCACCTGTTCACCTC	731
QY	545	ctcgtctgcaaatcttggaggtcggaagccctgtgtcaactaggaagcctagaagggctag	604
Db	732	CATTGCTGCAATTTAGAGAGTTGGACCATCATTTGATCTAGAGCACTGAAAGGACTAGG	791
QY	605	agaggtgtcacaatcacagccaagatgcaatgtgctctcaatctggccctccctctga	664
Db	792	AGAGGCTTTACATTTCCAGGCATGACATGTCGTCTCTTGGGCTCCCTCTTCA	851
QY	665	aagaagcaagctctctgagatctcaatctcagaagcaacaacttggagcagtgattctct	724
Db	852	AAGAAGCAAACTTCTTAGCATTTCTATTCAGAGAGACACACTTGGGAGATTAATTTCTCT	911
QY	725	tcctctcttgagtaatttggctctactatagaattgacctatgctcttatctcttg	784
Db	912	TCTCTTTCTGGAAATATTTGCTACTATTAAGAAATGGACTATGCTCTACTTTTGTG	971
QY	785	catgttgaaatctcgtgttatctttatgtactcctgaatgaatgatacacaagaac	844
Db	972	TACTATTGGAATATTTTGGTTCTTTTGTGTGACTGCTGTTAGTAGACACACCAAA	1033
QY	845	tcaacaagaacaatccccaatgaaagagatlatctcttcalcatlaaaaaatcagct	904
Db	1032	ACACAGAGATTTCCCATTAATGAAAGAAATACATTTCTTCACTTAATTAAGAAATCAGCT	1091
QY	905	ctcttcaagaagtcagtgccgtggaactatgctggaatacctctgcaacttgggctat	964
Db	1092	TTCTTCACAGAAAGTCAGTGCCTGGTGATCCATTTTAAAAATCCCTCCACCTTGGGCTAT	1155
QY	965	tgtctgtcaacttctcttaacaactgaccttttctaactgttgtaacctatgtctac	1024
Db	1152	CGTAGTTCACACTTTTCTTACAACTGAGATTTTATACTTATTTACATTAATGTGCTAC	1211
QY	1025	ttacaatgaaggaagtcctaaagttcaaalatcaagaagaatgggtttlatctgcagctcc	1084
Db	1212	TTAATATGAGAGAGATTCCTAAGGTTCAATGTTCAAGAGAAAGGGTATTATCTTCATTTGCC	1277
QY	1085	ttatttgggtgtgtgtatgtatgatccgtctgggtcaagcgtcgtgacaatttgaaggc	1144
Db	1272	TTATTTAGGGCTTGTGTATGTATGATCTGCTGTGGCAAGCTGTCGACAAATTTAAGGGC	1331
QY	1145	aagatgaatttccaactcgtgtgttgtaagaagtttttlaagcctatagagatgtg	1204
Db	1332	AAANATGGAATTTTTCACATTATGTGTCCGCAAAATTTTAAAGCTTATAGAAATGATAGG	1391
QY	1205	aactgtgatalctcgtgtgcgcgagattatagcgtgtatlatctctgtgtgtgc	1266
Db	1392	ACCTGACATATTTCTGTGATGCTGCTCATTTGGCTGTGATTAATTTCTTGGCCCTTGC	1451
QY	1285	attcctaaccatcaacaacccctgggaagcctttgtctctctcgattttagatcaaca	1324
Db	1452	TTTCTTAATATATCAACACACACTGGGAGGCTTTTGTCTTCTTGGAATTAAGATCAACCA	1511
QY	1335	tcctgaaactgtcctctcgtatgtgttatctccctgggatacaaaatacccttccac	1384
Db	1512	TCTGGAATTTGCTCTCTTCGTATGCTGTGATATCTCTCGGAGATACAAATACATTTTGCCAC	1577
QY	1385	tattccttgaaatgtgtggcccatcatctgcagaagatcttaacccttgagaaactatgtg	1444

Db 1572 TATTCAGAGATGTTGGGCCGCTATGCTAAAGTCTGACCCCTGATACTGTTGG 1631
 Qy 1445 agaattgcaaacatgttttctgcacgtctgctgcatacaatgatttggtgcattctt 1504
 Db 1632 AGAATGGCAAAACCGTGTATATATATGCTGCTGCTATTAATGTTTTTGTCATTTCTT 1691
 Qy 1505 cacacttcgcacaaaggtgaagtgcaaacatggccatcgatgatacccaagaccag 1564
 Db 1692 TACACTTTTCGCCAAGAGTGAACAAACTGGGCTCTCAATATATCACTGATGACACAG 1751
 Qy 1565 aaactgaaggaaaccaataaataatcctgctctataatgatacttctgttattatgtaa 1624
 Db 1752 ACACGTGAAGCAACCAATTAATTAATCCCTGCTATTAATGATTTTATTTATCATGTAA 1811
 Qy 1625 cctaaagtgcttgatttattatgtgtaagaacatctatatacaagaataaattgtac 1664
 Db 1812 CCTCAAAAGTGGCTT-----CGTATTTGTATAGCATTCTATGT-----CTTTTTTT 1857
 Qy 1685 taagaanaattggttgaatttgaagcttgtaactgaatgaatgcaatgagtgccatat 1744
 Db 1858 AATGTACTTGTATATTAATTTTAAAGCCTATATATCATGAATATCACTAGTTCAGAA 1917
 Qy 1745 aagcaaaattagctatttttaattataataacccgttggctggaacttaacaattcagg 1804
 Db 1918 TAATAAATGAAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1977
 Qy 1805 tcaatattctgctgcaagtcgaacccaacaaatgaaggagattctatta-ttataag 1863
 Db 1978 TCATATATCGTCCCTGCTGATGCGGCAACATGAATGAAGACATGCTGTGATTTTAAAG 2037
 Qy 1864 accatacctaagaagatgaactgaatagaacccctcctacacttgccttgaatgaagtg 1923
 Db 2038 GCCATCTTAAGGAATAGAGGTGAACAGACCTTCTGATACCTTGTCTTAATTAATAG 2097
 Qy 1924 ataataattcctgaagcttgtaaacacttgcttggtaacacttcccaaaaaattatt 1983
 Db 2098 ATGATTAATTTCTCAAGTACTGATTAACACCTGTTGTTCTCACTTCCCTCATTAATA 2152
 Qy 1984 tgcataagaacatccctgaatgaatgtagtctcaaacctttagacttccacaggaagtcag 2043
 Db 2153 -ATTGTAGGCTCTCTGACACTTGAACCTTAACCTTTAGCATCTGTGTGAGTGGCAT 2211
 Qy 2044 ccaactgatacatcagcctgcaactcactgaaggaagcaatgcacagcgagctgcacaa 2103
 Db 2212 CCACTGATATATTTGCGCTGCGCAACTGTGACAGGAGAGTGTGCCAGGACGCTGCCAAG 2271
 Qy 2104 tgcctcctctcctgctcagggagagtgccagactt-----aggcagcatcaaga 2156
 Db 2272 CACTCCCTCCCTGCTTCAAGGCTCAGAGTCCAGGCTTTATCAGAGGAGCATCCCAAGC 2331
 Qy 2157 ccaaggtgacgcgcgaagccttggaggtattcttccctc-aggagtgtaattggttg 2214
 Db 2332 CCAAGAGCAAGTGTGACTCTTCGCTGCTGCTTCTCTGAGGGGTATCAATAGTGTAG 2391
 Qy 2215 atgaagcctgagccaacaggaagcgcgatccacagatgatttccatgacacccctc- 2273
 Db 2392 ATTAAGCCCGAGTAGGCAAGAGAGATGACATCGCATGTGTCTTGAATACATCTCA 2451
 Qy 2274 tcccttcccttccacagcaactgagatttgctgagcatgtaacctgcaaaagaa 2329
 Db 2452 AACTTTCCTTCCACAGACAGAGAAATATTGGCTGAGTCAACCTCAAAAGAA 2507

RESULT 4
 HSA387747
 LOCUS HSA387747 2512 bp mRNA linear PRI 10-DEC-1999
 DEFINITION Homo sapiens mRNA for stalin.
 ACCESSION AJ387747
 VERSION AJ387747.1 GI:6562532
 KEYWORDS stalin.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2512)
 Verheijen, F.W., Verbeek, F., Aula, N., Beers, C.E., Havelaar, A.C.,
 Jooisse, M., Peltonen, L., Aula, P., Galjaard, H., van der Spek, P.J. and
 Mancini, G.M.
 A new gene, encoding an anion transporter, is mutated in sialic
 acid storage diseases
 Nat. Genet. 23 (4), 462-465 (1999)
 20047778
 2 (bases 1 to 2512)
 Verheijen, F.W.
 Direct Submission
 Submitted (03-JUN-1999) F.W. Verheijen, Erasmus University, Dept.
 Clinical Genetics, P.O. Box 1738, 3000 DR Rotterdam, NETHERLANDS
 Location/Qualifiers
 1..2512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 271..1758
 /codon_start=1
 /product="stalin"
 /protein_id="CAB62540.1"
 /db_xref="GI:6562533"
 /translation="MRSPPVRLDARNDSESTDTPLPGAPRAEAPVCCSARYNLA
 LAPEGFIVYALRVNLSVALVDVDSNTLEDRKSCPEHSPAPKRVHNOTGKRYO
 WDAENQWILGSPFFGYITQIPDGVAKIGKMLIGLSTAVLTLEPTIADLG
 VGPLIYLRALBEGEVTPPRAHMMSSNAPREKRLSISAGQITVLSPLSG
 ICTYVNMVTVFFETGTFIHWLIMTWISDTPONHKRISHREYILISLNQLS
 OKSPVWPILKSLPLMAIYAVHFSYMTFTVLTLLPTVMKELIRNVENGFLSLP
 YLSWLCMLISGQADNLRKAMNFSTLCYVRIFSLTGMIGAPVAVLAAGFIGDYSIA
 VAFITSTLIGSGSGFSINHDIAPSVAGILGLTGNFATPGWGVYIAKSLTPD
 NNYGEMOTFYIAINVFALINVFALIFETLFAFAGEVQNMALNDHHGRH"
 BASE COUNT 612 a 585 c 566 g 749 t
 ORIGIN
 Query Match 55.3%; Score 1572.4; DB 9; Length 2512;
 Best Local Similarity 82.1%; Pred. No. 0;
 Matches 1917; Conservative 0; Mismatches 386; Indels 33; Gaps 8;
 Qy 5 ggaggcgagggtcctcgagcgtccgctggaactctcttcccgaggaagcagttgccc 64
 Db 194 GGCCCGGGGGCTTCGCGCTTCGGCGCGGCGCTTCTCTGCGCAGGTGCGAGTACAC 253
 Qy 65 gtatgctccctgaaggtactaagttcccggttcttggaattgccccggagcgagcga 124
 Db 254 CT-GCTCAGCTAGCGCTCATAGAGTCTCGGTTGAGACCTGCGGAACGATGGCGA 311
 Qy 125 ggaaggtcgaacgcgacacacgcctcctcagcgccgcgcgcgcgcgcgcgcgcgcgc 184
 Db 312 GGAAGAGCAGGACCGCAGCGCTCTTTCACCGGGCGCCCGCCAGCGGCGAATCGTCCAGT 371
 Qy 185 atgctgctgctgctgttaacaactagacttctgctcttcttggttctcgttctccta 244
 Db 372 GTGCTGCTCTGCTGCTTACCAATTTAGCAATTTTGGCTTTTGGTTCTTCAATGTGTA 431
 Qy 245 ttcaattcgggtgaatcttgagcgttgacacagtggaatggttgattcaacaacaactgc 304
 Db 432 TGCAATTACGTGTGATGTGAGTGTGGTTGATGATGATGATGATGATGATGATGATGAT 491
 Qy 305 caaagataatagaacgtcccaagctgttcagagacatcctgcctcccaataaagttctca 364
 Db 492 AGAAGATATAGAACTTCCAAAGCGCTTCCAGAGCATTTGCTCCCTTAATAATTCATCA 551
 Qy 365 caacccaacggttaaaaagtaacggttggtgaatgaagaactcaagagatgattctcgatc 424
 Db 552 TAATCAAAACGGTAAAGATACCAATGAGATGAGAAATCAAGATGATGATGATGATGATG 611
 Qy 425 tttttctatggtatacatcaacaataatcctctgagagatagttgccaggaagtg 484
 Db 612 CTTTCTTATGCTATCATCATCACAGATTCTCGAGGATATGTTGCCAGCAAAATAGG 671

OY	485	ggggaagcgtgtgtcgtgagattcgggaatccttctgaagatattcctacccgtttacccc	544
Db	672	GGGGAATATCTGCTAGAGATTGGGATCCCTGGCACTGCTGTCACACCCTGTTTACCTCC	731
OY	545	ccctgcgcagattcgcgaatcggagccctctgtgcaccacgaaggcaactagaagggctagg	604
Db	732	CATTGCTGCAGATTATAGGATTGGACCACCTCATGTGACTCAGAGCACTRGAAGGACTAGG	791
OY	605	agaaggtgtcaacatatacgaacgaatgacatgcatgtgtctcaatgagctcccccctctga	664
Db	792	AGAGGGTGTATCATTTTCCACGCCATGATGATGATGATGTGTCTCTTGAGGCTCCCTCTTGA	851
OY	665	aagaagcaagcttctgtagtatttccatgcaaggagcaaaccttggagcaagtattcctct	724
Db	852	AAGAAGCAAACTTCTTGTGCAATTTCCATATGAGAGACACAGCTTGGGACAGTATTTCTCT	911
OY	725	tcaccttctcggagtaattgtgcacataagaatgagcttaagtctctatcttctcttgg	784
Db	912	TCCCTCTTTCGGAATTAATTTGCTACTATATGAATTGGACTTATGTCCTTCTACTTTTGGG	971
OY	785	catgttggaaatcatctgtgttatttctatgaaatcgttagttagtatacacaagaaac	844
Db	972	TACTATTGGAAATATTGTTGGTTCTTTTGGAGATCTGGTTAGTTAGACACACACAA	1031
OY	845	tcacaagacaatacctccctcctatgaaaaagatatacttccatcaatnaaaatcagct	904
Db	1032	ACACAAAGAAATTTCCATTTATGAAAAAGGAAATCATTTCTTCATTAAGAAATCACCT	1091
OY	905	ctcttcacagaagtcagtgccgtgagtcactatgctgaataacgcgaacttggctat	964
Db	1092	TTCTTTCACAGAGTCACTGCGGGGTACCCATTTTAAATCCCTGCACTTGGGCTAT	1151
OY	965	tgctgttcacatatttcttcaaacgcggaactttaaactgttgcacttaatcttgcac	1024
Db	1152	CGTAGTTGCACACTTCTTCTTACACACGAGACTTTTATACTTATATGACTATATTCCTAC	1211
OY	1025	ttacaatgaaggaagtcctaaagttcaatatcaagaagaatggttctatctgcagctcc	1084
Db	1212	TTATATGTAAGAGAGATCTCTAAGGTTCAATGTTCAAGAGAAATGGGTTTATCTTCAATTCGCC	1271
OY	1085	ttatttagtgttggtagtatagtatgcgtgcgtgcgaagctgcgtgcgaacttaataagagc	1144
Db	1272	TTATTTAGGCTCTTGGTATGTAATGTAATCCTGCTGCTCAAGCTGCTGACAAATTTAAGGCT	1331
OY	1145	aagaatgaatttccaactctgtgggtctcgaagagttttaagccttataaggaatattgg	1204
Db	1332	AAATGGAATTTTTCAACTTATGTGTGCGAAGATTTTACCCTTATATGAAATGATTTGG	1391
OY	1205	acctgagatattccctggtctgcgcgaagattatagagctgtgatatctccctgcgtgtgc	1264
Db	1392	ACCTGCAGTATTCCTGTGTACTCTGTGGCTTCATATGGCTGTGATTAATCTTGGCCGTGTC	1451
OY	1265	attcctatacattatacaaacacctggagagctttgtctctcttggaatttagatacaaca	1324
Db	1452	TTTTCTTACATATATCAACAACACTGGGAGGCTTTTCTTGTGATTTTATGATCAATCA	1511
OY	1335	tcctgaacttgcctccttcgtagtctgtatattccctgcgggcatcaacaataaccttggcac	1384
Db	1512	TCGTGGATATTTGCTCTCTTGATGCTGTGTATCTCTCGGGGATCACAAATACATTTGGCAC	1571
OY	1385	tattcctggaatgatttggcgccatctatgcgaagaatcttaacccctgagaacaacttgg	1444
Db	1572	TATTTCCAGGAATGGTTGGCCCGCTCATGTGGTAATAATGTGACCCCGAATAACAGTGGG	1631
OY	1445	agaatgcaaacctgtttctgcacatgcgtctgctatcaatgtatatttgtgtgcatttctt	1504
Db	1632	AGATATGCAAAACCTGTTCTATATTCCTGCTGCTATTAATGTTTGGGTCATTTTCTT	1691
OY	1505	caacattctgcgaagaagtgaatgtaaaaaactgggcatcaagtatgatacccaagacaag	1564
Db	1692	TACACTATTTGCCAAAGGTAAAGTATCAAAACTGGGCTCTCAATGATCACTATGGACACAG	1751
OY	1565	aaacttgaaggaaccaataataatcctgtctcattcaatgtaattcttgtttctatgttaa	1624

Db	Accession	Sequence	Length	Source
Db	1732	ACACGGAAGGACCAATTAATATATCCCTGCTCTATTAATGTAATTTTATTTATTAATGTA	1811	
Qy	1625	ccataaagtcgccccttgataatttaaigtgtaagcaacatcatalaacaagaataaatgttac	1684	
Db	1812	CCTCAAGGCCCT-----CTGTATTGGGTAGACATTCATGT-----CTTTTTTT	1857	
Qy	1685	tagaaaatgtgttagatttgtaagacctgttaacatgaataatgtcactagttgccatat	1744	
Db	1858	AATTTACTTGTATTAGATTTTAAAGGCCCTTAATCAATGAATATACACTAGTTGCCAATA	1917	
Qy	1745	aagcaaatgtactattttaattatattiaaccggtttgtcgtgaacttaacattcaagg	1804	
Db	1918	TAAATAATGAACCTGTGTTTAAATTATGTAATATATATATATATATATATATATATATAT	1977	
Qy	1805	tcacatctctgctgcaagtccagcaaccccaataggggagttcattta-ttataag	1863	
Db	1978	TCACATTAACCTGCTCTAGTCGCGCAACATGATAGACACAGTCTGTGATTTTAAAG	2037	
Qy	1864	accatacctaagaagatgagctgtaaatagaccctctatlaacctgtgcttaataagttg	1923	
Db	2038	GCCATTAATAAGGAAATGAGTGAAGAACAGACCTCTGATACCTTTGCTTAATTAACATAG	2097	
Qy	1924	ataataattctcaggtctgtttaaacatcgttttttgtaacctctcccaaaattatt	1983	
Db	2098	ATGATTAATTTCTAGGTACTGATTAACACCTGTTGTTGTTCACTTTCCTATTAATAA----	2152	
Qy	1984	tgtaacgaacatccctgacatgttagtgctcctaactttaagcctccagggagctggag	2043	
Db	2153	-ATTGTCAAGCTCTCTGTACACTTATGACCTTAACCTTAATGACATCTCTGTGGAGCTGCCAT	2211	
Qy	2044	ccactgtatcattcagcctgagcaactcactgaaggaagacatgcccagcagctgcaca	2103	
Db	2212	CCACTGTAAATTTGGCGCGGCACTGGACACTGAGGAGAGTGTCGCCAGGACACTGCCAAG	2271	
Qy	2104	tgccccctctctgcttgaaggagacgtgccagcaact-----aggcagcatccaaga	2156	
Db	2272	CACCTCCCTCCCGGCTTACAGGGTCAGAGGCGCCACGTTTATCAGAGGACGATCCAAAGC	2331	
Qy	2157	ccagggttcagcgaagacctgttgagacgtatctccct--gggacgttaattgtg	2214	
Db	2332	CCAGAGCCAGTGTCCAGCTTTCGGCTGTGGCTTTCTCTGTAGAGGGGCTATCAATGTGTAG	2391	
Qy	2215	atgaagccctgagccaacagggacagcgcgatccacagctcagtggtttccatgacacctc-	2273	
Db	2392	ATTAAGCCCTGAGTATGAGCAAGACAGGTGATGCCACTGCTATAGGTGATTCATCTCA	2451	
Qy	2274	tccttcctctccagcacactgtgagttatgctctgtgcatgttaacctgcacaaagaa	2329	
Db	2452	AACCTTCCCTCCAGCAGCAGAGAAATATTTGGCTGGCATGCAACCTGCAAAAGAAA	2507	
RESULT	5			
LOCUS	AX207624	2930 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0156598.			
ACCESSION	AX207624			
VERSION	AX207624.1	GI:15422329		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	Helicobacter, C.G.			
JOURNAL	Methods for preventing or attenuating pathogenic conditions by			
FEATURES	using the gbs-toxin (cna101) receptor as a vaccine			
source	Patent: WO 0156598-A 1 09-AUG-2001;			
	VANDERBILT UNIVERSITY (US)			
	Location/Qualifiers			
	1..2930			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			

CDS
263..1873
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60358.1"
/db_xref="GI:15422330"
/translation="MAGAMTPPPRPVQPARPGFGISGRSLICQVASTPAHVGMRS
PVLDLNDDEESTDTPILLGAPRAEAIVCCSARNLILAFFGFIVYALRVNIS
VALVDVNSVTTLEDNRTSKPEHSPIVKHNPQKXOMPELQVILGSPFYGY
ITIQIPGVYASKIGGKMLGFIIGLAVTLPTPIADIGVPLVRLLEGISEGY
TFPAMHAMSSMAPLERSKLISYVAGLGVISLPSGIIICYYMMTVYVFFRG
IGTFWLLIMWISDPQKRIKSHKEKILISLRQDSQSVPRVILKSLPLMA
IVVHFSYNTFTYLLLEPTLMEKILRFNVQENGFLSLPTLSMCMILSQADN
LRKMNFSYLCVRIFSLIGMIGPVLVAAIGDYSLAVFLISTLGGSCSG
FSINHDIAPSYAGILIGTINTEATIGPVPVIAKSLTPDNTVGEMQVETIAAII
VFGAIFETLPAKSEVQWMLNDHGHNR"

BASE COUNT 729 a 698 c 681 g 822 t
ORIGIN

Query Match 55.2%; Score 1570.8; DB 6; Length 2930;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 1916; Conservative 0; Mismatches 387; Indels 33; Gaps 8;

QY 5 ggggggggggctcgccggtcccgctggagctctcttcccgaggaagttgcgc 64
DB 309 GGGCGGGGGGCTTCGCGCTCGGGCGGGCGGCGCTTCCTCTCCAGGTGGCGAGTAC 368
QY 65 gtagctccctaaagcatcatgaagtcgccggttcgagtaagcccgagcgagcga 124
DB 369 CT--GCTACAGTACGCGTCAAGAGTCTCCGTTTCGAGACTGGCCCGAGACATGGCGA 426
QY 125 ggaaggtcggagccgacacacgcctcctcgagcgcccgcgaggcggaacccgctcagt 184
DB 427 GGAAGACGACGACGCGACGCGCTTCTACCGGGCGGCCCGACGAGCGAGCTCCAGT 486
QY 185 atgtctcgtcgtctgttacaacctaagcattttgctcttcttggcttcgttcccta 244
DB 487 GTGCTGCTCTGCTCGTTACCACTTGAATTTGGCTTTTGGTTCTTCATTTAGTGA 546
QY 245 ttcatcaggggtgagctcgagcggttcgacatgtagcatgtgattcaacaacatcgc 304
DB 547 TGCATTACGCTGATCTGAGTGTGGCTTACGTATGTGTAGATTCATTAATACACTTT 606
QY 305 caaagataatagaacgtccctactagatgtagagcatctgctcccaataaagttctca 364
DB 607 AGAAGATAATAGAACTTCCAAGSGGTGCCAGACATTTGCTCCATTAAGAATGATCA 666
QY 365 caaccacaaggttaaaaagtaacgcgtgtgtagtgaagaactcaagaatgattcgcgac 424
DB 667 TAATCAAAACGGGTAAAGATACCAATGGATGCAAGAACTCAAGGATGGAATTCGCGTTC 726
QY 425 ttcttcatggctacatcatcacacaatctcctggaggaatgttccagcagaagttg 484
DB 727 CTTTTTTATATGCTACATCATCACAGATTTCCCTGGAGGATATTTTCCAGCAAAATAGG 786
QY 485 ggggaagcgtgtcgagatcgggattctgtcacagctatctccaccctgttactcc 544
DB 787 GGGGAAATATCTGCTAGATTTGGGATCTTGGCAGTGTCTCCACCTGTGTACATCC 846
QY 545 cctcgcgagcagatttgggaatcgagccctgttgcactaaggcactaaagggctagg 604
DB 847 CATTGCTGCAAGATTAGAGATTGGACCATCTCATTTGTACTAGACACTTAAGACATGAG 906
QY 605 agaaggttcacatcatccagcatgcatgcatgtgttcttcaatgggctccccccttga 664
DB 907 AGAGGGGTTCATATTTCCAGCATGATGCCATATGTGTTCTTTGGGCTCCCTCTTTGA 966
QY 665 aagaagaacgtcttgatattcaatataggaagcaacaacttgggaacagtagttctct 724
DB 967 AAGAAGCAAACTCTTAGCATTTGCTATGCAAGACGACACTTGGGACAGATTAATTTCT 1026
QY 725 tcctctcttcgtgaatattgtactatatagaattggaattatgtcttattcttgg 784
|||||

DB 1027 TCCCTTTTCGGAAATATTTGCTACTATATGAAATGGACCTTATGCTTTTCTTTTGG 1086
QY 785 catgttggaaatcatctggttatttattatgacatctgtagttagatcaccaagaac 844
DB 1087 TACTATTGGATATTTTGGTTTCTTTTGGATGTGTTAGTATGATACACACACAAA 1146
QY 845 tcacaagaacatctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 904
DB 1147 ACACAAGAGATTTTCCCATATGAAGAGATATACATCTTTCATCATTAAGAAATACGCT 1206
QY 905 ccttccagaagtcagtgccgtagtactatgctgaatcaactgcacttgggctat 964
DB 1207 TTTCTTCAAGACAGTCAAGTGGCGTGGACCATTTTAAATCCCTGGCACTTTGGGCTAT 1266
QY 965 tgcgttcgacatcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1024
DB 1267 CGTAGTTGCACACTTTTCTTACAACTGACCTTTTATACCTTTATGACATTTAGCTTAC 1326
QY 1025 ttacatgaaggaagtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1084
DB 1327 TTATATGAGAGATCTCTAAGGTTCATATGTTCAAGAGATGGGTTTTATCTTCATTTGCC 1386
QY 1085 ttatttagtctgtgtatgtagtgcctgctgggttcaagctgctgaacatlaaaggc 1144
DB 1387 TTAATTTAGGCTCTTGTATGATGATCCCTGCTGCAAGCTGCTACAAATTTAAAGGC 1446
QY 1145 aagaatgaattttcaactcgtgtgttcgaagaagttttagccttataaggaatgtagt 1204
DB 1447 AAATGGAATTTTCAACTTTATGTGTTCCAGAAATTTTATAGCTTTATAGATGATGG 1506
QY 1205 aactgcgaattccctgtgtcccgaggaattatagctgtgatbatctccttggcgttgc 1264
DB 1507 ACCTGCAATATCTCGTACTGCTGCTGCTCATTTGCTGATATTTCTTTGGCCGCTTGC 1566
QY 1265 attcctacacatcaacaaccccgaggagcctttgtctctcctcctcctcctcctcctc 1324
DB 1567 TTTCTTACATATATCAACAACACTGGGAGGCTTTTGTCTTCTGATTTAGCATCAACA 1626
QY 1325 tctgacatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1384
DB 1627 TCTGGATTTGCTCTCTCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1686
QY 1385 tctcctcgaatgattgggcccacatctgcagaagcttaccctcctcctcctcctcctcct 1444
DB 1687 TATTCAGGAATGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1746
QY 1445 agaattgcaaacgcttctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1504
DB 1747 AGAATGGCAAAACCGTGTATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1806
QY 1505 cacactatcgcacaaggtgaagtgaagcaaacctggccatcagtgatcaccaaggaacag 1564
DB 1807 TACACTTTTGGCCAAAGGTGAAGTACAAACTGGGCTCTCAATGATACCATGAGACAG 1666
QY 1565 aaactgaaggaacccaataaatacctcgtctctatataagtgatcttgtttatcatgtaa 1624
DB 1867 ACACTGAAGGAACCAATTAATATCCGCTCATTAATGTAATTTTATTAATATCATGTAA 1926
QY 1625 cctaaagtgctgttgaatatttaagtgaagaacatcctatatacaagaataaattgac 1684
DB 1927 CCTCAAAAGTGCCTT-----CTGATTTGTGTAACCATTTATGTA-----CTTTT 1972
QY 1685 tagaanaatctgttagatttgaagcctgtgaacatcagtaacatcagctagtcacat 1744
DB 1973 AATTGTACTGTATTTATGATTTTAAAGCCCTATATATCATGAAATATCATAGTGGCCAGAA 2032
QY 1745 aagcaaaatagctatttattatattatcaaccgtttgcgtgaacttaacatcagg 1804
DB 2033 TAATTAATAATGACGTGTTTAATTATGAATAATATGTAACCTAGACTTCTACTTTAGT 2092
QY 1805 tcacatatctcgtctgcaagtcagcaaccacaataagggaggttctattta-cttataag 1863
DB 2093 TCACATPACCTGCTGCTAGTGGGCAACATGAAGTAGACAGTCTTGTGATTTTATGAG 2152
|||||

OY	1864	accactactaaagatgtagctgaatatgagccctcttaaccttgaacttggcttaattgaagtg	1923
Db	2153	GCACACTCTAAAGGGAGATAGCTGGAACAGACACTCCTGATACCTTTGGCTTAATTAACATG	2212
OY	1924	ataaatactctcaagtgcttgtaaacatctggttttggtaacacttccctcctaaanaattatt	1983
Db	2213	ATGATAAATCTTCAGGTACTGATGAATAAACACTGTTGTTGTCACTTCTTCATCAATAAA----	2267
OY	1984	tgctcatgaacaatccctgaacatgtagtgccttaacttagctcttccaaagggctggcag	2043
Db	2268	-ATTTCTAGCTCTCTCTAACACTTAGACCTCAAACTTTAAGCATCTCTGTGGAGCTGCCAT	2326
OY	2044	ccactgtatcatctacgctgcgaacttcaacttgaagaaagcatggccagcagctgccaca	2103
Db	2327	CCACGTGTAATATTTCGCTGGCCTGGCAACTGGACCTGAGGGAGTGTGCCACAGACACTGCCAAG	2386
OY	2104	tgctccctctctgctgcttcaaggagcagtgcccaagact-----aggcagcatccaaga	2156
Db	2387	CACTCCTCCCTCCGCTTCAGGCTCAGAGTGCCTCCACGTTTATCAAGAGGACGATCCAAAC	2446
OY	2157	ccaggtgtaacgcgcgaagcgtttgacgcgttatcttccct--ggggcgcttaattgtg	2214
Db	2447	CCAGAGCCAGATGTCACACTTTCGGCTGGTGGCTTTCTCTCTGAGGGGGCTATCAATGTGTAG	2506
OY	2215	atgaagccctgagccacaagagacagcgatcacagtcacagtcatalggtttccatgcacacct-	2273
Db	2507	ATTAAGCCCTGATGTCGCAAGACAGCATGATGATCCACTGCTATGGCTTGATTCATCTCA	2566
OY	2274	tcaccttccctcccaagcaacatgtagtcttggcttgcagcatgtaactgtgaanaagaa	2329
Db	2567	AACTTTCCTTCCACGACACAGAGATATGTGCTGTGCATGCAACTGCAAAAGAA	2622

RESULT	6
AF244577	
LOCUS	2930 bp
DEFINITION	Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, linear
ACCESSION	AF244577
VERSION	AF244577.1 GI:9719373
PRI	06-AUG-2000
CDS	complete cds.

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 2930)	Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P., Carter,C.E., Shi,E., Venkov,C., Yakes,M.F., Page,D.L. and Hellerqvist,C.G.	Identification of a novel membrane protein from mammalian cells that interacts with the anti-pathoangiogenic compound CMI01	Unpublished	2 (bases 1 to 2930)	Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R. and Hellerqvist,C.G.	Direct Submission	Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School of Medicine, 3rddeperce, Nashville, TN 37232-0146, USA
Location/Qualifiers	1..2930						

```

gene      1. . 2930
          /gene="HP59"
          263. . 1873
          /gene="HP59"
          /note="similar to Na+ inorganic phosphate transporter"
          /codon_start=1
          /product="membrane glycoprotein HP59"
          /protein_id="AAF97769.1"
          /db_xref="GI:9719374"
          /translation="MAAGAMPMPRPVOPAPRGSGLSRRSLICOVASTPAHVGMNR
          PYRDLARNDEGEESTDRTPILPGAPRAEAPACCCSRKYLALITAEFGFTVALRVNCS
          VALVQVDSNTLTLENNRTSKACPEHSAPLTKHHNNTGKQWDAETQGMILGSEFYGG

```

BASE COUNT
ORIGIN

729 a 698 c 681 g 822 t

VFCAAGFTFLFAKGGVQWMLNDHGGHH"

	Query Match	Similarity	85.2%	Score	1570.8	DB	9	Length	2930	
	Best Local	Similarity	82.0%	Pred.	No	0				
	Matches	1916	Conservative	0	Mismatches	387	Indels	33	Gaps	8
QY	5	gggggaggggagccttcgagcgctccgcgtctgagagctctcttccgcgaggaagagttgagc	64							
Db	309	GGCCCCGGGGGCTGCGGGCTGTGCGGCGCGGCGCTCCTTCTGTGCGCAAGGTGGGAGTACAC	368							
OY	65	gtagctccctgaagagcatcatgaagttcccggtttcgaacttagcccgagcgagcgaga	124							
Db	369	CT--GCTCAACGATGGCGTCATGATGAGGTCTCCGGTTTGGAACCTGGCGCCGGAACGATGGGGA	426							
OY	125	ggaagggtccgagaccgacacccgcgtctgcaagcgccccgcgagggcgaggaaccgcctcaat	184							
Db	427	GGAGAGCACAGGACCGCACGCCCTCTCTACCGGGCCGCCACAGGGCGGAGAGCGCTCCAGT	486							
OY	185	atgctgctctgctcgttcaaacctagcaattttgtcccttlltggltttcttcgtctcta	244							
Db	487	GTGCGCTCTGCTGCTGTACCAACTGTAGCAATTTTGGCCCTTTTGGTTTCTTTCATTGTGTGA	546							
OY	245	ttcattacggggtgaatctcgagcgctgtgcaatagtgaaactgtggtgttcaaacacacgyc	304							
Db	547	TGCATTACGTGTGAATCTGAGTGTGCGTGTATGTGATGTGATGATTCAAATACACTTTT	606							
OY	305	caaaagataaagaaagtcctacgaagtgcgcaagagatctcgtcccaataaaagttcttca	364							
Db	607	AGAAATATATATAAATCTCCAGCGGTGTGCCAAGAGATTCTGCTCCCATTAAGATTATCA	666							
OY	365	caaccaaacgggtlaaaaaagtlaccggtgtagagtcagaaactcaaggaatgattcttcgagtc	424							
Db	667	TAAATCAAAACGGGTGAAGGTATACATGGGATGTGCAGAACTCAAGAGATGGATTCTCGGTTC	726							
OY	425	ttttctcagtgctacatcatcaacaaattcccttgaaagatatgttcgacgaagaagtg	484							
Db	727	CTTTTATATGGCTATCATCATCATCAACAATTCCTGGAGATATGTGTGCAGCAAAATATGG	786							
OY	485	ggggaagctgtgtcctagagatctcgagatctttgctacagcatcttcacaccttcaatcc	544							
Db	787	GGGGAAATATGCTGCTAGAGATTTTGGATCTCTTGGCAGCTGCTGTGCTCACCCGTTCACGCC	846							
OY	545	ccctgcctcagaatctcgaagtcggaagccctgtgtgcaactcaaggagcactagaagagctagg	604							
Db	847	CATTCTCTCAGATTTTAGAGATGGAGCACTCATTTTACTACAGAGCACTAGAAGGACTAGG	906							
OY	605	agaagggtcacatatccagcccatgacatgacatgtggtgtctcaatgaggtccccccttga	664							
Db	907	AGAGGGTGTACATTTTCCAGCCATGCAATGCCATGTGGTCTTCTTGGGCTCCCCCTCTTGA	966							
OY	665	aagaagcaagctctgaatatttcatalctacatgacgaggaacaaacttggagacagtagttctct	724							
Db	967	AAGAGCAAAACTTCTTACCATTTCTGTATGAGGACACAGCTTGGGAGACAGTAATTTCTCT	1026							
OY	725	tcctcttctcgtgagaaatttgcctaactatgaatggagactatgctcttacttctcttgg	784							
Db	1027	TCTCTTTCTGGAAATATTTTGTCTACTATATATGATGTGACTATATGCTTCTACTTTTGTGG	1086							
OY	785	catgtctggaatcatctgttatttatttlaagatcgtcgtagttatgataaacacgaac	844							
Db	1087	TACTTTGGAAATATTTTGGTTCTTTTGTGGATGTGGTTAGTTAGTGAACACACACAAADA	1146							
OY	845	tcacaagaacatcacctccctatgaaagagatatcttcttcatactaaataacacagct	904							

Db	1147	ACACAAAGAAATTTCCATTATGTAAGAAAGAAATACATTTCTTATCATTTAAGAAATACACT	1201
Oy	905	ctcttcacagaagtctagtcgcgtggaataactgctgaatactacgtccacttggctat	964
Db	1207	TTCTTCCACAGAAAGTCACTGCGGCGGTGATCCATTTTAAATCCCTGCCACTTGGGGCTAT	1266
Oy	965	tgctgttcacatcttctctacaacgtggacttttaacttgcttgacttatggctac	1024
Db	1267	CGTAGTTGCACACTTTTCTTACACATGGACCTTTTACTTTATTTGACATTTATTTGCTTAC	1326
Oy	1025	ttacatgaagaaagtcctaaggttcaataatcaagaagaatgggtttttatctgcagtc	1084
Db	1337	TTATATGAAAGAGATCCTAAGTGTTCATATGTTCAAGAGATGGGTTTTATCTCATTTGCC	1386
Oy	1085	ttattagtgctgttggttatgatatgatactctgtcgtgcgaactgctgcacaatlaaaggc	1144
Db	1387	TTATATTTAGGCGCTTGGTATGATGATGATCCGTCGTCACAGCTGTGCAATTTTAAGGCG	1446
Oy	1145	aagaatgaattttcaactctgtcgggttcgaagaagatttttagccctatagaagatattg	1204
Db	1447	AAAAATGGAATTTTCAACTTTATGTGTTCGCAAAATTTTAACTTTATATAGCAATATTTGG	1506
Oy	1205	accgtgcatactccgtgttgcgcgaagattlaagcgttgatlaatlccttgcgtgtgc	1264
Db	1507	ACGTGCAGTATTCCTGGTACTGCTGGCTTCATTTGGCTGTGATTTATCTTTGGCGGTGCC	1566
Oy	1265	atccctataccatatacaaacacctggaggactttgctctctcttgatttagcatcaaca	1324
Db	1567	TTTTCTTACTATATATCAACAACACTGGGAGAGCTTTTGCTCTTGAGTTATGATCAACCA	1626
Oy	1325	tcgtgaactgtccctctgcgtatgctgtatctccgtggcatcaaaaactcttgcac	1384
Db	1627	TTCTGGATATTTGCTCTTGATCTGATCTGTGATTCCTCCGGGCATCAACAATTTATTTGCCAC	1686
Oy	1385	tattcccggaatgatattggcccaatctatgcgaagctctaacccctgaaacaactattgg	1444
Db	1687	TATTCAGGAATGGTTGGGCGCCGCTCATTTGCTAAAGTGTGAGCCCTGATTAACACTGTTGG	1746
Oy	1445	agaatggcaaacctgtttctctgcatgcgtcgtcatcaatgatttggctgcattctc	1504
Db	1747	AGAAATGGCAAAACCGTGTTCATATTTGCTGCTGCTATATATGATTTTGGTGCCATTTTCTT	1806
Oy	1505	cacactatctgcaaaagggtgaagttgaacaacgtggccactcagtgatgcgaagaacag	1564
Db	1807	TTACACTTTTGGCCAAAGGTGAATGACAAAACGTGGGCTCAATGATGACCAATGACACAG	1866
Oy	1565	aaactcgaaggaaaccaaataaatacctcgtcctatlaatglatctcttggatcatgtaa	1624
Db	1867	ACACTGAGAAGACCAATTAATATATCTGCGCTCTATTAATGATTTTATTTATCATGTAA	1926
Oy	1625	cctaagaagtccttggatatttlaatgtglaagcaactatatacaagataaaatgttac	1684
Db	1927	CCCTCAAGTGCCTT-----CTGATTTGTTGATAGCAATTCATGT-----CTTTTTTTT	1972
Oy	1685	tagaaaaattgttgatttatttgaagcgttgatcaatcagaatgtcactagtgtccatat	1744
Db	1973	AATGTGCTCTGTATTTAGATTTTAAAGCCCTAATATATGAAATATACATCACTAGTCCAGAA	2032
Oy	1745	aagcaaaattagcatattttaactatctaaacccgttttgcgtgaacttaacttcaggg	1804
Db	2033	TAAATAAATTAAGACTGCTGTTTAATTAATTAATTAATATATGTAAGCTAGGACTTTCATTTAGGT	2092
Oy	1805	tcacatactcgtgtcgaagtcagcgaacccaacaatagggaggtctatatta-tttataag	1863
Db	2093	TCACATATCTGCTGCTAGTCGGGCAACATGAAATGAGACAGTTCTGTGATTTTTTTAGG	2152
Oy	1864	accatacctaagaagatgagctgaaatagaccctctataactgtgcttaattgaagtgg	1923
Db	2153	GCCATCACTAAAGGAATGAGCTGAAACACACCTCTGATACCTTTGCTTAATTTAACTAG	2212
Oy	1924	ataataactcagagctctgtttaaacaattgttttttgaacacttcccaaaaatatc	1983
Db	2213	ATGATATTTCTCAGGTAAGTGAATAAACACTGTTGTTGTCATTTCCGCATAAAAA-----	2267

OY	1984	tgctaaagcaatccctcaacatgtagtgctcaaaactttgcctctccaaagagctgagag	2043
Db	2268	-ATTGCTAGCTCTCTGTGACACTTGAACCTCAAACTTTTAGCATCTCTGTGGAGCTGCCAT	2326
OY	2044	ccacgtatcatcagcctcgagcaacttcaactgaaggaaagcatgcccagtcagctgcaca	2103
Db	2327	CCACGTGTTAATTGTGGCCCTGAGCACTGACACTGAGGGAGATGTGCCAGGACACTGCCAAG	2386
OY	2104	tgtccctctctgctctcaaggagacagtgcacgaactt-----agcagcatccaaga	2156
Db	2387	CACCTCCCTCCCTGGCTTCAGGGCTCAGAGAGGCCAGGAGTTATTCAGAGGACGATCCAAAC	2446
OY	2157	ccaaggtcaagcgccaagctcttgagacggtatcttccct--gggctgttaatgtgtg	2214
Db	2447	CCAGAGCCACAGTCCACACTCTTGCGGTGGTGCCTTCTCTTGAGGGGCTATCAATGTGTAG	2506
OY	2215	atgaagcccttgaagccaagaaggaacagcgagatcccaagtcagaggttccatgcaacctc-	2273
Db	2507	ATTAAGCCCTTGAGTAGGCAAGACAGTAGATCCACTGCTATGTGTCCTATCAATCCCTCA	2566
OY	2274	tcccttcccttcccaagcaacatgtagtatttgctctgcatgttaactgtgcaaaagaa	2329
Db	2567	AACTTTCCCTTCCACGACAGAGGAAATATTGGCTGGCATGCACAACCTGCCAAAAGAA	2622

	RESULT	7		
LOCUS	BC020961			
DEFINITION	Homo sapiens, solute carrier family 17 (anion/sugar transporter), member 5, clone MGC:8885 IMAGE:3847279, mRNA, complete cds.	3329 bp	mRNA	linear
ACCESSION	BC020961			PRI 22-JAN-2002
VERSION	BC020961.1			
KEYWORDS	GI:18088702			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	1 (bases 1 to 3329)			
TITLE	Strausberg,R.			
JOURNAL	Direct Submission Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	Contact: MGC help desk Email: cgapbs-remai.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94307 Web site: http://www.sbgc.stanford.edu Contact: (Dickson,Mark) mcdgpaxil@stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.			
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov</p> <p>Series: IRAK.Plate: 20 Row: i Column: 2</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 691265.</p> <p>Location/Qualifiers</p>			

```

/organism="Homo sapiens"
/db_xref="LocustID:26503"
/db_xref="taxon:9606"
/clone_image="MGC:8885 IMAGE:3847279"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NH_MGC_65"
/lab_host="DH10B"
/notes="Vector: PCMW-SPORT6"

```

CDS
125. .1612
/codon_start=1
/product="solute carrier family 17 (anion/sugar transporter), member 5"
/protein_id="AAH20961.1"
/db_xref="gi:1808703"
/translation="MRSPVRLARNDSESTDRPFLPGARAEAPVCSARYNLAI
LAFGEFLVALRVNLAVLADWDSNTLEDNRSKAPRHSAPIKVHNQTKKYO
WDATGGWILGSFEGYIIIOIGVYASKIGSKMLGFLGTRAVLLEFPIADLG
VGPLVRLALEGLSEGTFFPAMHMSWMAKPLESKLISYGAQLGTVISLPLSG
LICVMNTVYFEFFGIFWFLMWLVDTPQKRIKISHYENFVLESLRQDLS
QKSVWPEILKSLPMAIVAHFSTNMFYLLDLPYMEILRFENQENGFSLP
YLGSMWILSGAADNLRAKNFSTLCVRIFSLIGIMGAFLVAGFVAGFCVSLA
VAPLISITLLGFCSSGFSINHLIDAPVAGILGINTFPTIGMWPVAKSLTLP
MTVGEWQVFIFIAAINFEAIFELPFRKGVQWMLNDHGHNR"

BASE COUNT 863 a 718 c 727 g 1021 t
ORIGIN

Query Match 55.2% Score 1570.8; DB 9; Length 3329;
Best Local Similarity 82.0%; Pred. No. 0; Mismatches 387; Indels 33; Gaps 8;
Matches 1916; Conservative 0;

QY 5 ggggagggggggttcgcgcgtccgcgtgagctctctcttccgcgagcaggttgcgc 64
DB 48 GGGCCGGGGGCTTCGGGCGTGCGGCGCGCGCTCCCTTCCTCCAGAGTGGCGAGTAC 107
QY 65 gtagctccctgaagcagcatatgaagtcccggttccgaactaagcccgagcagcgga 124
DB 108 CT--GCTACAGTAGGCGTCATGAGTCTCCGGTTCGAGACTGCGCGGAAAGCGGCA 165
QY 125 ggaaggtcgcgagccgcaacgcgctctgagcgcccgccggcgagcggaaccgcctcagt 184
DB 166 GGAAGAGCAGACGCCGACGCCCTCTTCTACCGGCGCGCCCGACGGCGGAGCCGCTCCAGT 225
QY 185 atgctgctctgctgttacaacctagcatttctgcttcttctgttctctctctcta 244
DB 226 GTGCGCTGCTGCTGCTTACAACTTAGCAATTTGGCCTTTTGGTTCTTCATTGTGTA 285
QY 245 ttcatlaagcgtggaatcgcgagcgttgcacatagtgacatggtgattcaacacaaatgc 304
DB 286 TGCATTACCTGTGAATCTGAGTGTGCGGTAGTAGTATGATTCATAAATACAACTTT 345
QY 305 caaagataataagacgtccctacagtgtagagcattctgctccataaagctcttca 364
DB 346 ACAAGATATATACAATCTCCAAAGCATGTCCAGACATTTGCTCCATTAAGATTCAATCA 405
QY 365 caaccaaacggttaaaaagtlacccggttggatgcagaaactcaaggatgattctcgaat 424
DB 406 TAATCAAAACGGGTAAAGAGTACCAATGGGATGAGAACTCAAGATGATTCGCGTTC 465
QY 425 tttttctatgctacatcatcaacaatctccctggagagatagtgtgcagcagaagaatgc 484
DB 466 CTTTTTTATATGGCTACATCATCATCAGACAGATTCCTGAGGATATGTTGCCAGAAAAATAG 525
QY 485 ggggaagcgtgttgcagagttcggagatcttgcacagctatctcaaccctgttaacctc 544
DB 526 GGGGAAAAAGCTGCTAGAGATTGGGATCTTGGCACGTGCTGCTCACCCTTTTACTTCC 585
QY 545 cctgcgtcagaatttcgagatcgcgagcccttgtagcactcaagcgagcactagaaggatcag 604
DB 586 CATGTGCTCAGATTAGGAGTGTGACACTCATTTGACTCAGAGCACTAAGAGAGACTAGG 645
QY 605 agagaggtgtaacatatcagccatgcagatgcctgtgtcttcaatgagctccctccctgga 664
DB 646 AAGAGGTGTATTAATTTCCAGCCATCATGCCATGTGTGTTCTTGGGCTCCCTCTTCTGA 705
QY 665 aagaagcagaactctgagatctcatatgacagagcacaactgtgagacagttatctct 724
DB 706 AAGAAAGCAAACTTCTTAGCATTTCAATGCAAGAGCACAGCTTGGAGACAGTATTTCTCT 765
QY 725 tccctcttcctgagtaattgtctactatagtaattggaacttattctctatctcttgg 784

DB 766 TCCCTCTTCTGGAATAATTTGCTACTATATGAAATTCGACTTATGTCCTACTTTTTGG 825
QY 785 catgttggaaatcaccatcgtgttattttagatcctgctagttagtgataaccgcgaagac 844
DB 826 TACATATTGGAATATTTTGGTTCTTTTGGATCGGTAGTATGAGACACACACAAAA 885
QY 845 tcaagaacaaactccctccatgaagaagagatattcttcaatcaatlaaaaatcagct 904
DB 886 ACACAGAGAAATTTCCCATTTATGAAAAGAAATACATCTTTTATCATCTTAAGAAATCAGCT 945
QY 905 cctctcacaagaatcagtgccgttggatacctatgctgaaataactgcccacttgggctat 964
DB 946 TTTCTACAGAGTACGTCCGCGGTACCCATTTTAAATCCCTGCCCTTTGGCGTAT 1005
QY 965 tctgctgcacatttcttcaaacctggaacttllatacttgttgaccttatgtacctac 1024
DB 1006 CGTAGTTCGACACTTTTCTTACAACTGGCACTTTTATACCTTATGACATTTATGCTAC 1065
QY 1025 ttacatgaagaagctccctgaagttcaatatlaagaagaaatgggttllttagctcgaagctcc 1084
DB 1066 TTTATATGAAGAGATCCTTAGGTTCAATGTTCAAGAGAAATGGGTTTATATCTTCAATGCC 1125
QY 1085 ttatttaggttggtttagtatagtatccctgcgcggtcgaagctgcagacaatllaaggcc 1144
DB 1126 TTAATTAGGCTCTTGGTATGATGATCCTGCTGCTGCAAGCTGCTGACAAATTTAAGGCG 1185
QY 1145 aagatggaatttccaactcgtggttcgaagaagttttagacctataggatgattgg 1204
DB 1186 AAAATGGAATTTTCAACTTATGTTGTCGAGAAATTTTATACCTTATGAGATGATTTGG 1245
QY 1205 acctgcgaatctccgtgtgcgcgaagattataagacgtttagtattctctgctgctgtgc 1264
DB 1246 ACCTGCAGATTTCCGTGATGCTGCTGCTGCTTCAATTTGGCTGATTTATCTTTGGCGGTTC 1305
QY 1265 attcctbaacatalcaacaacaccctgggaagcttctgctctcttgcgatttagacalcaacca 1324
DB 1306 TTTCTTAATATATATCAACAACTGAGAGGCTTTTGTCTCTCGATTATAGATCAACCA 1365
QY 1325 tctggaacatgctcctcttgtagtgcgtgatatctcctgggcatacaaaactcttggcac 1384
DB 1366 TCTGATATATTGCTCTGCTATGCTGATGCTCTCTGCGGCATCAACAAATTTGGCCAC 1425
QY 1385 tatctcgaatgattgggcccacatcgtccagaagcttcaacctgagaaacacattgg 1444
DB 1426 TATTCAGGAATATGTTGGCCCGGTCAATTCATAAAGTCTGACCCCTGATTAACACTGTGG 1485
QY 1445 agaaltggcaaacgttcttgcacgcgtgctgctatcaatglatattgtygcacttctt 1504
DB 1486 ACAATGCGCAAAACCGTCTTATATGCTGCTCTATTTATGATTTTGGTGGCATTTTCTT 1545
QY 1505 caactattcgcgaagaagtgaagctgcaaaactgggcccactcagtgataccaagaagacag 1564
DB 1546 TACACTATTTCGCCAAAGGAGTGAAGTCAAAACGCGCTCTCAATGATCACCAATGAGACAG 1605
QY 1565 aaactggaagcaacaataataatcctgctcatlaaatglatcttctgttcaatgtaa 1624
DB 1606 ACACGGAAGAGCAATTAATATATCTGCTCTATTAATATGATTTTATTTATTTATTCATGTA 1665
QY 1625 cctaaagtyccttgcataattlaatgtytgaagcaactcatatacaagaataaattgtac 1684
DB 1666 COTCAAAAGGCTT-----CTGTATTTGTATGACATTCATATGT-----CTTTTATTT 1711
QY 1685 tagaaaatgtgttagatttgaagccttgaatcaatgaaatgtcactagttgcataat 1744
DB 1712 AATTTACTTGTATTAATTTTAAAGGCTATTAATCAATGAAATATACATGATTTGACAGAA 1771
QY 1745 aagcaaatatgatacttctlaattatlaaacccgttgcgtgagcaactacaataccagg 1804
DB 1772 TATATTAATGAACCTGTGTTTAAATTTATGATTAATATGATGACTAGAGACTTTCTTAGCT 1831
QY 1805 tcaacatcgtgtcgaagtcagcgaaccacaataagggaggttcatatta-cttataag 1863
DB 1832 TCACATACCTGCTGCTAGTGGGCAACATGAAGTAGGACAGTCTGTTGATTTTGG 1891

QY	1864	acattractaagagatagctgaaatagaccctctctactctgtgctaatatagyg	1923
Db	1892	gcacatcctaaaggagatagctgaaacagacctcctgattacctgttaataaacgag	1951
QY	1924	ataataatctcaggtctgtttaaacatctgttttltgacacctctcccaaaatalt	1983
Db	1952	atgatatatttcaggtactgatataaacacctgtgtgtgttcaacttccctatataaa	2006
QY	1964	tgtcatcagcaatccctctacatgtgagtctcctaaacttaagcctccacaggaagctgacag	2043
Db	2007	-attgtcagcctcctcctacaccttaagacctcaaaattgagatctctgtggagcgccat	2065
QY	2044	ccactgtatcatctacgctgcgaacctctacatgaaggaagcatccacagcaagtccaca	2103
Db	2066	ccactgtattatttttcgtccctgcgcacactgacactgaaggagagtgctgccacagcactccacag	2125
QY	2104	tgtccctctctcgtgcttaagggaagtgcccaagcact-----aggcagcatccaaaga	2156
Db	2126	cacttccctccctcctgctttaggggtcagaagctgcccacagctttatcacagggcagacatccacagc	2185
QY	2157	ccaggtctaacgcccgaagctcttgagcaggtatctccct--ggggcggttaatgtfgyg	2214
Db	2186	ccagagccagatgtgcactcttgcgtgtgcttcttccctgtagggggctatcatatgtgtag	2245
QY	2215	atgaagcccttagcacaacaggaacagcgcagatccacagtcatagttctcatcacacct-	2273
Db	2246	ataaagcccttagatgtggcagaagcagatgagatccacgtctatgtctgattcatatcttca	2305
QY	2274	tcccttccctcccaagacaactgtgagtattgtcctggaatgttaacctgcgaagaaga	2329
Db	2306	aactttcccttccacgcacagaggaattgtgtgtgctgacatggcaacctgcgaataaagaa	2361

RESULT	8	
LOCUS	AK025880	
DEFINITION	AK025880 Homo sapiens cDNA: FLJ22227 fis, clone HRC01782, highly similar to HSA387747 Homo sapiens mRNA for stalin.	1587 bp mRNA linear PRI 29-SEP-2000
ACCESSION	AK025880	
VERSION	AK025880.1	GI:10438530
KEYWORDS	Oligo capping; fis (full insert sequence).	
SOURCE	Homo sapiens primary human renal epithelial cells cDNA to mRNA, clone_1lp:HRC clone:HRC01782.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (sites)	
TITLE	Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.	
REFERENCE	NEDO human cDNA sequencing project	
AUTHORS	Unpublished (2000)	
TITLE	2 (bases 1 to 1587)	
REFERENCE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.	
AUTHORS	Direct Submission	
TITLE	Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sunio	
JOURNAL	Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)	
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
FEATURES	Location/Qualifiers	
source	1..1587	
	/organism="Homo sapiens"	

[illegible]

|||||
Db 844 TGATACCTTGGCTTAATTAACAGATGATTAATTCAGGTACTGATTAACACCTGTGT 903
QY 1959 tgaacaccttcctcaaaaattatttgcatacgaacatccctgacatgtgcttcaaac 2018
Db 904 TGTTCACCTTCCATTAATA-----ATTGTACGCCCTCTGTACCTTAAGACCTCAAC 957
QY 2019 ttaagacctccagagctgagcgaactgtatcatctcagccctggcgaactcaactgag 2078
Db 958 TTTAGCATCTCTGTGAGCTGCGCATCTGATTAATTCGCTGGCAACTGGACATGAG 1017
QY 2079 gaagacatgccagcagcagctgcaacatgtccctctctggtcttcaaggagtgccca 2138
Db 1018 GGAGTGTGCCAGGAGCAGTCCCAAGACATCCCTCGCTTCCAGAGGTGAGAGTCCAG 1077
QY 2139 ctt-----aggcagcatcccaagcagaggtcagcgccaaagcttggacggtattct 2191
Db 1078 CGTTATATCAGAGGACATCCCAAGCCAGCAGTGTGACTCTTGCGCTGTGCTT 1137
QY 2192 cccct--ggagctgttaattgtgtgatagaagccctgagccaacaaggagcagcgatca 2249
Db 1138 CCTCTAGGGGGCTATCAATGTGTAGATAAGCCCTAGTAGGCAAGACAGTAGATCCA 1197
QY 2250 cagtcagtgttccatgcaacctc-tccctccctccccaagcacacatggaatgtgctg 2308
Db 1198 CTGCTATGGCTGTGATACATCTCAAACTTCCCTTCCAGCAGAGGAATATTGCTG 1257
QY 2309 gcatgtaacctgcaaaaanaa 2329
Db 1258 GCATGCAACCTCGAAAAA 1278

RESULT 9

AK026921 2006 bp mRNA linear PRI 29-SEP-2000
LOCUS AK026921
DEFINITION Homo sapiens CDNA: FLJ23268 fls, clone COL08932, highly similar to
HSA387747 Homo sapiens mRNA for slalin.
ACCESSION AK026921
VERSION AK026921.1 GI:10439893
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens colon CDNA to mRNA, clone_lib:COL clone:COL08932.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hiki,J., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEBD human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2006)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-ADU-2000) to the DBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center,
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEBD human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'-3' end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES

Source 1..2006
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL08932"
/clone_lib="COL"

/tissue_type="colon"
/note="cloning vector pME18SFL3"
misc_feature 1..2006
/note="highly similar to HSA387747 Homo sapiens mRNA for
slalin"

Query Match 21.2%; Score 602.8; DB 9; Length 2006;
Best Local Similarity 78.3%; Pred. No. 8e-127;
Matches 822; Conservative 0; Mismatches 197; Indels 31; Gaps 7;

BASE COUNT 559 a 418 c 420 g 609 t
ORIGIN
QY 1291 gaagcttggctcttcttgattagatcaaacatctgagcatgtctctctgtagctg 1350
Db 1 GAGGCTTTTGGCTCTTGGATTAGCATTAACCATTTGGATTTGCTCTCTGATGCTG 60
QY 1351 gtaattcctgggacatcaaaaatatacttggcactatctcctggaatgtggtccatca 1410
Db 61 GATTCCTCGGGCATCACAAATATATTTGGCCATATTCAGGAATGTTGGGCCGCTCA 120
QY 1411 ttgcagaagcttacccttgagaacacattggagaatggaactgtttctgcacag 1470
Db 121 TTGCTAAAAGCTGACCCCTGATMACCTGTGGAAATGGCAAAACGTTCTATATG 180
QY 1471 ctgcgtatcaatgtatattgtgtgcatcttcttcacactatctgcgaaggtgtagtgc 1530
Db 181 CTGCTCTATTATATGTTTGGTGGCCATTTCCTTACACTATTGGCCANAGTGAAATG 240
QY 1531 aaaactgggacatcagtgatcaaccaagaacacagaacactgaagaaaccaataatcc 1590
Db 241 AAAACTGGGCTGCATGATGATCACCATGACAGACAGACATGAAGAACCAATAATATCC 300
QY 1591 tgtctctatatagtatcttctgtttatcatgttaacctaaagtgccttggattta 1650
Db 301 TGGCTCTATTATATGTTTGGTGGCCATTTCCTTACACTATTGGCCANAGTGAAATG 355
QY 1651 gtgtgaagcatctatatacaagaataatgttgaataatgttggattgtagctg 1710
Db 356 GTGTAAAGCTTCTATGT-----CTTTTAAATTTGACTTGTATTAATTTTAAAG 406
QY 1711 gcttgaatcatgaaatgtcactagttgcatataagaacaaatagctatlttaattat 1770
Db 407 GCTATTAATCATGAATATATACATGAGTGGCAGAAATTAATGAATGATGTTAATAT 466
QY 1771 tattaaccggttctgctggaacttaacatcagggtacatatctgctgctgcaagca 1830
Db 467 GAATATATATGTAAGCTTAGGACTTCTACTTTAGGTTCCACATACCTGCTATGTCGGCA 526
QY 1831 accacaatgagggaggttctatatta-ttataagaccataccaaagatgagctgag 1889
Db 527 ACATGAAGTAGGACACTTCTGTGATTTTAAAGGCCATCAATGAAGATGAGCTGA 586
QY 1890 tagaacctctataaccttctgcttaataagtgtaataaatctcaaggtctgtttaa 1949
Db 587 CAGACTCCTGATACCTTTGGCTTAATTAACATAGATGAATATTCAGGACTGATAAAC 646
QY 1950 atctgttttctgtaacacttccctcaaaaattatttgcatacgaatccctgacatg 2009
Db 647 ACCTGTTGTTGTTACCTTCCATTAATA-----ATTGTACGCTCTCTGACACTAG 700
QY 2010 gctcaaaattagacctccagagctgagcagcagcagctgcatctcagctggaa 2069
Db 701 ACCTCAAACTTAGCATCTCTGTGGAGCTGCCATCAGTATTAATTTGGCTGGCAACT 760
QY 2070 tcaactgaggaagcatgccagcagcagctgcaacatgtccctctcctgctcagga 2129
Db 761 GGACTGAGGGAGGTGGCCAGGACATGCCAAGCACTCCCTCCCTGAGGGTACAG 820
QY 2130 tgcacagcactt-----aggcagcatcccaagcagaggtgcaagcagcttggac 2182
Db 821 AGTGCCACGAGTTTATCAGAGGAGCATCAAGCCAGACAGAGTGTGACATCTTGGCT 880

QY	2250	cagtcataagtttcattgcacccctc-tcccttccttcctccacgacacacgtgagatattgctg	2308
	1		
Db	43854	CGGCTATGCTGTGTTGATCATCTCTCAAACTTTCCTTCACGACACAGAGAAATATGCTG	43913
QY	2309	gcattgaaccttcataaaga	2329
	1		
Db	43914	GCATCAACCTGCAAAAGAA	43934

RESULT	11
LOCUS	HSJ397H23
DEFINITION	Human DNA sequence from clone RP3-397H23 on chromosome 6q12-14.1,
ACCESSION	AL121972
VERSION	AL121972
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE	1 (bases 1 to 113202)
JOURNAL	Kay, M.
COMMENT	Direct Submission Submitted (17-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk requests: clone.requests@sanger.ac.uk On Mar 12, 2001 this sequence version replaced gi:12832031.. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6 RP3-397H23 is from the library RPrCt-3 constructed by the group of Pleier de Jongh. For further details see http://www.choir1.org/bacpac/home.htm VECTOR: pCYPAC2 This sequence is the entire insert of clone RP3-397H23. Location/Qualifiers 1..113202 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /map="q12-14.1" /clone="RP3-397H23" /clone_lib="RPrCt-3" 1..58 /note="AluU/FLAM repeat: matches 5..62 of consensus" 260..570 /note="AluX repeat: matches 1..305 of consensus" 674..707 /note="MER20 repeat: matches 29..62 of consensus" 708..880 /note="AluSP repeat: matches 120..292 of consensus" 961..1095

repeat_region	/note="AluSg1 repeat: matches 1. .135 of consensus" 1105. .1428
repeat_region	/note="AluSx repeat: matches 1. .306 of consensus" 1432. .1585
repeat_region	/note="MER0 repeat: matches 60. .218 of consensus" 1826. .2123
misc_feature	/note="Alu1 repeat: matches 1. .295 of consensus" complement(2112. .2118)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data" complement(2120. .2127)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data" complement(2129. .2147)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data" complement(2149. .2160)
repeat_region	/note="Single clone region. assembly confirmed by restriction digest data" 2150. .2458
misc_feature	/note="AluSx repeat: matches 1. .306 of consensus" complement(2161. .2164)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data" complement(2168. .2263)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data" 2468. .2774
repeat_region	/note="AluSg1 repeat: matches 1. .308 of consensus" 2795. .2932
repeat_region	/note="FLAM_C repeat: matches 1. .118 of consensus" 3028. .3325
repeat_region	/note="Alu1 repeat: matches 1. .296 of consensus" 3355. .3644
repeat_region	/note="AluSg repeat: matches 1. .287 of consensus" 3775. .4074
repeat_region	/note="AluSx repeat: matches 1. .300 of consensus" 4096. .4386
repeat_region	/note="Alu10 repeat: matches 1. .298 of consensus" 4490. .4793
repeat_region	/note="Alu1b repeat: matches 1. .306 of consensus" 5117. .5253
repeat_region	/note="FLAM_C repeat: matches 1. .132 of consensus" 5501. .5814
repeat_region	/note="AluSg repeat: matches 1. .308 of consensus" 6097. .6225
repeat_region	/note="FLAM_A repeat: matches 1. .131 of consensus" 6296. .6479
repeat_region	/note="Alu1b repeat: matches 129. .312 of consensus" 6545. .6830
repeat_region	/note="Alu10 repeat: matches 42. .306 of consensus" 6965. .7270
repeat_region	/note="AluSg repeat: matches 1. .306 of consensus" 7308. .7602
repeat_region	/note="Alu1b repeat: matches 1. .295 of consensus" 7737. .8035
repeat_region	/note="AluSg repeat: matches 1. .299 of consensus" 8184. .8488
repeat_region	/note="AluSx repeat: matches 1. .299 of consensus" 8917. .9176
repeat_region	/note="AluSg1 repeat: matches 1. .293 of consensus" 9212. .9377
repeat_region	/note="L1M4 repeat: matches 6005. .6177 of consensus" 9490. .9785
repeat_region	/note="AluSx repeat: matches 1. .296 of consensus" 10352. .10431
repeat_region	/note="L2 repeat: matches 2616. .2702 of consensus" 10432. .10734
repeat_region	/note="AluSx repeat: matches 1. .301 of consensus" 10735. .10778
repeat_region	/note="L2 repeat: matches 2575. .2616 of consensus" 10846. .10985
repeat_region	/note="MIR repeat: matches 53. .200 of consensus"

repeat_region 11023..11140 /note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region 12005..12079 /note="LIMD3 repeat: matches 7662..7737 of consensus"
repeat_region 12350..12647 /note="Alusx repeat: matches 1..297 of consensus"
repeat_region 12669..12916 /note="Alusx repeat: matches 1..293 of consensus"
repeat_region 13058..13351 /note="Alusx repeat: matches 1..296 of consensus"
repeat_region 13504..13657 /note="AlusC repeat: matches 152..306 of consensus"
repeat_region 13838..13947 /note="Alus repeat: matches 25..138 of consensus"
repeat_region 13948..14254 /note="AlusG repeat: matches 1..305 of consensus"
repeat_region 14255..14437 /note="Alusx repeat: matches 138..307 of consensus"
repeat_region 14482..14610 /note="MERSA repeat: matches 46..167 of consensus"
repeat_region 14614..14911 /note="AlusC repeat: matches 1..301 of consensus"
repeat_region 14934..15229 /note="Alus repeat: matches 1..297 of consensus"
repeat_region 15266..15556 /note="Alus repeat: matches 5..299 of consensus"
repeat_region 15578..15888 /note="Alusx repeat: matches 1..311 of consensus"
repeat_region 15965..16062 /note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region 16157..16439 /note="Alus repeat: matches 11..294 of consensus"
repeat_region 16618..16913 /note="Alus repeat: matches 1..296 of consensus"
repeat_region 16933..16998 /note="33 copies 2 mer tt 69% conserved"
repeat_region 17001..17104 /note="AlusG/x repeat: matches 196..307 of consensus"
repeat_region 17135..17428 /note="AlusC repeat: matches 1..293 of consensus"
repeat_region 17790..18090 /note="Alusx repeat: matches 1..301 of consensus"
repeat_region 19220..19547 /note="Alus repeat: matches 2..290 of consensus"
repeat_region 20724..21025 /note="Alus repeat: matches 1..304 of consensus"
repeat_region 21045..21082 /note="19 copies 2 mer gt 89% conserved"
repeat_region 21084..21260 /note="AlusG/x repeat: matches 134..310 of consensus"
repeat_region 21367..21670 /note="MIR repeat: matches 32..165 of consensus"
repeat_region 21819..21939 /note="Alus repeat: matches 1..306 of consensus"
repeat_region 22647..22943 /note="Alus repeat: matches 1..299 of consensus"
repeat_region 23011..23316 /note="Alus repeat: matches 1..306 of consensus"
repeat_region 23399..23702 /note="Alus repeat: matches 1..304 of consensus"
repeat_region 24322..24622 /note="Alus repeat: matches 1..302 of consensus"
repeat_region 24680..24971 /note="AlusC repeat: matches 1..292 of consensus"
repeat_region 25178..25302 /note="AlusG/x repeat: matches 172..300 of consensus"
Query Match 16.9% Score 480.2; DB 9; Length 113202;
Best Local Similarity 75.7%; Pred. No. 1.2e-98;
Matches 697; Conservative 0; Mismatches 193; Indels 31; Gaps 7;
QY 1420 gtctacccttggaacactatgtggaatggaactgtttctcatcgtctgctgcta 1479
||||| | ||||||| ||||||| || ||||| || |||||||

Db 56637 GTCTTCCTTATATGAACACTGTTGGAGAAATGGCAAAACCGTGTCTATATGTCTGTCTGA 56578
QY 1480 tcaatgtattgtgtgcattttcttcaactatctgcgaagtggaagtgcaaaactgga 1539
Db 56577 TTATGTGTTTTGGTGGCATTGTTTACACTATTCGCCAAAGGTGAAGTACAAACGGG 56518
QY 1540 ccaatcagtgtaacaaagagacagaacatgaaggaacaaataaatctgtctcat 1599
Db 56517 CTCTCAATGATTCACCATGACACAGACACTGAAGGAACCAATTAATCTCTCTCTAT 56458
QY 1600 taatgtaactgtttatcatcgtlaacctaagaatgtcccttgatatttaagtgtgaaga 1659
Db 56457 TAATGTATTTTATATATCATATGTAACCTCAAAAGTGCCTT-----CTGATATGTGAAGCA 56403
QY 1660 attatataaataaataatgttactcagaanaattggtatagattgttaaggttgaat 1719
Db 56402 TTTTATGT-----CTTTTATATGTTGCTTGTATTAAGATTTTAAAGGCTTAAAT 56352
QY 1720 catgaatgtcactagtgcataataagaanaattagctatttaattatataacc 1779
Db 56351 CATGAATATACACTAGTGGCAGCAATTAATTAATGAACGTGTTTATTAATGAATATAT 56292
QY 1780 gtttgcctggaacttaacatcaaggtcatalctgtgctgcaagtcgaagcaaccacaat 1839
Db 56291 GTAAGCTAGGACTTCTACTTTAGGTTCACTACACTGCTCTAGTCGGGCAATGATGAAT 56232
QY 1840 agggaggttctatta-tttataagcaactaactaaagatgaagttgaatgaacctt 1898
Db 56231 AGGACAGTTCTGTTGATTTTAAAGGCAATCAATTAAGGGAATAGCTGAACAGACCTCC 56172
QY 1899 ctacacttgccttaattaaagtgagataataatctcaagttcgtttaaactgtttt 1958
Db 56171 TGATACCTTGTCTTAATTAATCAATGATGATTAATTCAGTACTGATTAATACACTGTGTG 56112
QY 1959 tgtacaacttcctcaaaaataatttgtcatcagaacatccctgacatgtaagtgtcaaac 2018
Db 56111 TGTTCACCTTCTCTCAATAAA-----ATTGTCAGCTCTCTGACACTTACCTCAAC 56058
QY 2019 ttaagctctcaagagcgtgagcagcagctgtatcttaagcctggaacttaactgaag 2078
Db 56057 TTAAAGCATCTCTGTGGAGCTGCACATCCAGTAAATTTGCTTGCAACTGGACACTGAGG 55998
QY 2079 gaagcatgcccagcagcctgacacatgctccctctctgcttcaaggaagtcgccaaga 2138
Db 55997 GGAGTGGCCCGACGAGCTGCCAAGACGACGCTCCCTGCTTCAAGGTGCAAGTGGCCAG 55938
QY 2139 ctt-----aggcagcatccaagcaggggtcagcgcgaagcttgaacgtattctt 2191
Db 55937 CGTTTATCAAGAGCAGCATCCCAAGCCAGCCAGTGTGACCTTCGGCTGTGCTCTT 55878
QY 2192 ccgctt-ggggctgttaatgtgtgagtgaaagccctgaagccaagaaggaagcagcgatcca 2249
Db 55877 CCTGTGAGGGGCTATCAATGATGTAGATAAGCCCTGATGAGCCAAAGCAGTGAATTCGA 55818
QY 2250 caatcatggtttccatcagccctc-tcccttcccttcccaagcagcatggaagtatgctg 2308
Db 55817 CTGCTATMTGTCGTGATACATCTCCCAAACTTCCCTTCCACGACAGAGAAATTTTGCGTG 55758
QY 2309 gaatgtaacctgcaaaaaga 2329
Db 55757 GCATGCAACCTGCAGAAAGAA 55737
RESULT 12
AC034271 149597 bp DNA linear HTG 12-MAY-2000
LOCUS Homo sapiens chromosome 6 clone RP11-598619 map 6, WORKING DRAFT
DEFINITION
SEQUENCE, 31 unordered pieces.
ACCESSION AC034271.2 GI:7770480
VERSION AC034271.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens


```
misc_feature      8985..10630
                    /note="assembly_fragment"
misc_feature      10731..13202
                    /note="assembly_fragment"
misc_feature      13303..15992
                    /note="assembly_fragment"
misc_feature      16093..18722
                    /note="assembly_fragment"
misc_feature      18823..21695
                    /note="assembly_fragment"
misc_feature      21796..25265
                    /note="assembly_fragment"
misc_feature      25366..28014
                    /note="assembly_fragment"
misc_feature      28115..31378
                    /note="assembly_fragment"
misc_feature      31479..34284
                    /note="assembly_fragment"
misc_feature      34385..37707
                    /note="assembly_fragment"
misc_feature      37808..41922
                    /note="assembly_fragment"
misc_feature      42023..45878
                    /note="assembly_fragment"
misc_feature      45979..49890
                    /note="assembly_fragment"
misc_feature      49991..54041
                    /note="assembly_fragment"
misc_feature      54142..59531
                    /note="assembly_fragment"
misc_feature      59632..65537
                    /note="assembly_fragment"
misc_feature      65638..71486
                    /note="assembly_fragment"
misc_feature      71587..76694
                    /note="assembly_fragment"
misc_feature      76795..82579
                    /note="assembly_fragment"
misc_feature      82680..90920
                    /note="assembly_fragment"
misc_feature      91021..99034
                    /note="assembly_fragment"
misc_feature      99135..110324
                    /note="assembly_fragment"
misc_feature      110425..119810
                    /note="assembly_fragment"

Query Match      16.9%: Score 480.2; DB 2; Length 149597;
Best Local Similarity 75.7%: Pred. No. 1.2e-98;
Matches 697; Conservative 0; Mismatches 193; Indels 31; Gaps 7;
```

```
Dh 75653 CATGAATATACACTAGTGGCCAGATAATATAATGAACGTGTATTAATATGATAATAT 75712
Qy 1780 gtttcggaaactatacatcaggtacatctgctcgaagtcaagcaaccacaat 1839
Dh 75713 GTAAGCTAGGACTTCTACTTTAGGTTCCATACCTGCTCTGTCGGCAACATGAACT 75772
Qy 1840 agggagttctatla-ttataagacatacctaaagaatgagctgaaatagaccct 1898
Dh 75773 AGGACACTTCTGTGATTTTAAAGGCCATACATAAAGGAATAGCTGAAGAACAGACTCC 75832
Qy 1899 ctataccttgcttaataaagtgagataatctcaggtctgttaaacatctgttt 1958
Dh 75833 TGAATACCTTGTGCTTAATTAACATAGATATATCTCAGGTACGTGATTAACACCTGTTGT 75892
Qy 1959 tgaacacctctcctcaaaaattattgttaacgaatccctcagatgtagtctcaaac 2018
Dh 75893 TGTTCACCTTCTCATTAATAA-----ATTGTACAGCTCTCTCAGCACTTAGACCTCAAC 75946
Qy 2019 tttagcctctcaagagctgagcagccatgatacattcagcctgagcaactcagag 2078
Dh 75947 TTTAGACATCTCTGTGGAGCTGCGCATCCAGTATTAATTGCGCTGGCACTGACAGAG 76006
Qy 2079 gaagcatgccagagcagctgacacatgtccctctctgctcaggaagcagtgccagca 2138
Dh 76007 GGAAGTGTGGCCAGGACAGCTGCCAAGCACTCCCTCGCTCAGGCTCAGAGTCCAG 76066
Qy 2139 ctt-----aggcagatccaagacagaggttagcagcgaagcttgagcgglatctt 2191
Dh 76067 CGTTTATCAGAGGCGACATCCAAAGCCAGAGCCAGTGCACATCTTGGCTGGCTTT 76126
Qy 2192 cccct--ggggctgttaatgtgtgaaagccctgagcccaagagcagcagatca 2249
Dh 76127 CCTGTGAGGGGCTATCAATGTAGATAAGCCCTGAGTGAAGAGCAGTATGATGATCA 76186
Qy 2250 caatcatggtttccatgacacctc-tccctccctcccaagcagcaggtttgctgt 2308
Dh 76187 CTGCTATGTCTGTATCATCTCAAACTTCCCTTCCAGCAGAGATATATGTGCTG 76246
Qy 2309 gcaatgtaactgcaaaagaa 2329
Dh 76247 GCATGCAACCTGCMAAAGAAA 76267

RESULT 13
AC025535/c 157749 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-206H23, WORKING DRAFT
DEFINITION
ACCESSION AC025535
VERSION AC025535.4 GI:7658462
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 157749)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 157749)
            Waterston,R.H.
            Direct Submission
            Submitted (10-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On/Apr 28, 2000 this sequence version replaced gi:7582731.

COMMENT ----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0206H23
```



```

Db 92414 GCATGCAACCTGCAAGAAA 92394
||||| |||||||
RESULT 14
AX118967
LOCUS AX118967 375 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 131 from Patent WO0129221.
ACCESSION AX118967
VERSION AX118967.1 GI:14035921
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 375)
TITLE Conklin,D.C. and Yee,D.P.
JOURNAL Proteins and polynucleotides encoding them
Patent: WO 0129221-A 131 26-APR-2001;
Zymogenetics, Inc. (US)
FEATURES
source
1..375
Location/Qualifiers
CDS
1..375
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:14035922"
/translation="MIGPAFLVAAGFIGDYSYLAVALFTSTLGGFCSSGFSINHL
DIAPSYAGILAGITNFATIPGWGPIAKSLPDMVTGEMQTFYIAAINVEGAIF
FTLFAKGEVQNMALNDHGHHRH"
BASE COUNT 87 a 87 c 79 g 122 t
ORIGIN
Query Match 11.0%; Score 314.2; DB 6; Length 375;
Best Local Similarity 89.9%; Pred. No. 4.1e-61;
Matches 337; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1197 atgattgacctgagatattcttcgttgcgagagattatagctgtgattcttcctg 1256
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 ATGATTGGACCTGCAGATTCTCGTGTAGCTGGCTGCTGATTCGATTCCTTTG 60
QY 1257 gcttgcattccccaacatacaaacaccctggaggcttcttccttcgtattagc 1316
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 GCCGTTCCTTCCATATATACACACACTGGAGGCTTTTCCTTCCTGGATTAGC 120
QY 1317 atcaacatctgacatctgctcctctgtaagctgtaattctcctggcacaataacc 1376
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 ATCAACCATCTGGATTGTGCTTCGTATGCTGTATTCCTCGTGGGCATACAAATACA 180
QY 1377 ttgcacattcttgtaagtgtggccatcatctgccaagaatcttaacctgagaac 1436
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 181 TTTTCCACTATTCAGAAATAGTTGGCCCGCTCATTCCTAAACTGACCCCTGATTAAC 240
QY 1437 acatctggagaatgcaaacgttttctgcatcgtcctctcactcaatgatttggacc 1496
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 241 ACTTTTGGAGAAATGGCAACCGTCTTATATATTCCTCTGCTATTAATGTTTTGGTCC 300
QY 1497 atttctcacaactatcgccaagtgtgaagtgcacaactggtggcaccatgataccaa 1556
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 301 ATTTTCTTTACATATTCGCCAAAGGTGAAGTACAAACAGGGGCTCATGATGATCACCAT 360
QY 1557 ggacacagaactga 1571
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 361 GGACACAGACACTGA 375

```

```

ACCESSION AB020527 Homo sapiens mRNA for Na/P04 cotransporter homolog.
VERSION AK024903
KEYWORDS oligo capping; fis (full insert sequence);
SOURCE Homo sapiens colon cDNA to mRNA, clone_11b:COL clone:COL01253.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
TITLE Kawabata,A., Hkiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
JOURNAL Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 3616)
REFERENCE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
AUTHORS Shibahara,T., Tanaka,T. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'-6 3'-end one pass sequencing: Departent of
Virology and human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1..3616
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="COL"
/clone_11b="COL"
/note="cloning vector pME18SFL3"
BASE COUNT 994 a 785 c 820 g 1017 t
ORIGIN
Query Match 9.5%; Score 269.6; DB 9; Length 3616;
Best Local Similarity 52.7%; Pred. No. 8.1e-51;
Matches 609; Conservative 0; Mismatches 544; Indels 3; Gaps 1;
misc_feature
1..3616
/note="highly similar to AB020527 Homo sapiens mRNA for
Na/P04 cotransporter homolog"
QY 390 tggatgacagaactcagaatgattctcgtatcttttctcattgatacatcaca 449
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 432 TGGAGTCCCTGAATCCAGGGAATCATCTCAGCTCCCTCAACATGAGCTCATTCCTGGCT 491
QY 450 caaatctcggagagatattgctcagcagaagtgggggagagctgttgcaggaattcgg 509
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 492 CCAATCCCAAGTGGCTATGCTGGAATATTTGGAGACCAGATATGCTGGTGGCTGCC 551
QY 510 atcttgcacagcatctcaacctgttcaacctccctgcgtcagaatttcggagatcoga 569
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 552 TTGCTTATTTCCATATTCCTGACCCCTCTTCATTCACATGCGACGATGAGGAGTGGCC 611
QY 570 gccctgttgcactcagcagactagaaggtctagaagaggtgttcaatatccagcaag 629
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 612 TTGCTCATTTGCTCCGAGATTGTAACAAGGATAGCCAGAGTATGATTAAGTGTGCTAG 671
QY 630 catgccaatggtctcattcagctccccccttgaagaagaagctctgagattatca 689
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 672 TATTCAATTTGGGTCAATGGGCTCCGCCACTGGAAGAGAGTCAACTCACACACATTGCT 731
QY 690 tatgcaggagacaaacttggagacagtagttctctctctctctcttctgaggaattgtcac 749
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 732 GGATCAGGGGTCAATGCTGGGGTCTTCATGTCTACTTGTCTGTGTGTCTCTGCCAG 791

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 13:52:14 ; Search time 592.65 Seconds
(without alignments)
8239.103 Million cell updates/sec

Title: US-09-776-865-3
Perfect score: 2844
Sequence: 1 cccggggggcggggggtctcg.....ataaaaaaaaaaaaaaa 2844

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	2844	100.0	2844	AAZ50876	Sheep GBS toxin re
2	2844	100.0	2844	AAD10326	Sheep group B beta
3	1572.4	55.3	2512	AAF55900	Human ASF coding s
4	1570.8	55.2	2930	AAZ50879	Full length human
5	1570.8	55.2	2930	AAD10325	Human group B beta
6	1567.6	55.1	2602	AAZ50875	Partial human GBS
7	1458.8	51.3	2712	AAK94876	Human full-length
8	1413	49.7	1485	AAZ50880	Human/Sheep consen
9	1305	45.9	1485	AAZ50881	Human/Sheep consen

10	1200	42.2	1488	22	AAI58115	Human polynucleoti
11	1169.8	41.1	2670	22	AAH79234	Human sodium depen
12	672.8	23.7	1975	22	AAH99626	Human protein enco
13	642.4	22.6	929	22	AAI59901	Human polynucleoti
14	594.8	20.9	853	22	AAK93901	Human cDNA clone r
15	553	19.4	838	22	AAK92364	Human cDNA 5'-end
16	456.2	16.0	1975	22	AAH99626	Human protein enco
17	341.8	12.0	798	23	AAH66219	DNA encoding novel
18	341.8	12.0	1066	23	AAH7186	DNA encoding novel
19	314.2	11.0	375	22	AAH52158	Human AFP protein
20	229.8	8.1	349	22	AAI13566	Human breast cance
21	226.6	8.0	1641	23	ABL18113	Drosophila melanog
22	226.6	8.0	1786	23	ABL07417	Drosophila melanog
23	223.8	7.9	3422	22	AAK52406	Human polynucleoti
24	199	7.0	264	22	AAI23435	Human breast cance
25	192.6	6.8	2291	23	ABL20393	Drosophila melanog
26	192	6.8	1880	23	ABL1735	Drosophila melanog
27	190.6	6.7	1620	23	ABL14775	Drosophila melanog
28	190.6	6.7	3620	23	ABL14774	Drosophila melanog
29	190.6	6.7	90104	22	ABL12402	Drosophila melanog
30	186.8	6.6	2889	22	AAI93801	Human polynucleoti
31	185.2	6.5	2716	17	AAI42064	Human brain Nat de
32	185.2	6.5	2716	19	AAV35503	Human sodium-11thl
33	164.2	5.8	2281	19	AAV57909	Human haemochromat
34	160	5.6	1943	23	ABL09181	Drosophila melanog
35	159.8	5.6	1562	21	AAAC4267	Arabidopsis thalia
36	155.6	5.5	1643	19	AAV43711	Human sodium-depen
37	154	5.4	1795	19	AAV57910	Human haemochromat
38	146	5.1	300	20	AAZ14190	Human gene express
39	142.2	5.0	1939	23	ABL03769	Drosophila melanog
40	140.4	4.9	4899	23	ABL03768	Drosophila melanog
41	133	4.7	1512	23	ABL15787	Drosophila melanog
42	132.2	4.6	2003	23	ABL08217	Drosophila melanog
43	129.4	4.5	2280	23	ABL05527	Drosophila melanog
44	126.6	4.5	1440	23	ABL12049	Drosophila melanog
45	126.6	4.5	3440	23	ABL12048	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAZ50876	AAZ50876 standard; cDNA; 2844 BP.
ID	AAZ50876 standard; cDNA; 2844 BP.
XX	XX
AC	AAZ50876;
XX	XX
DT	31-MAY-2000 (first entry)
XX	XX
DE	Sheep GBS toxin receptor (SP55) cDNA.
XX	XX
KW	Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55;
KW	pathological vascularisation; cancer metastases; anglogenesis;
KW	neovascularisation; reperfusion injury; scarring; keloid;
KW	chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW	endothelial cell proliferation; antibacterial; anticancer;
XX	anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
OS	Ovis sp.
XX	XX
XX	XX
FT	Key
FT	Location/Qualifiers
FT	CDS 84..1571
FT	/*tag= a
FT	/product= "Sheep GBS toxin receptor"
XX	XX
PN	WO200005375-A1.
XX	XX
PD	03-FEB-2000.
XX	XX
PF	22-JUL-1999; 99WO-US16676.
XX	XX
PR	22-JUL-1998; 98US-0093843.
XX	XX

PA (UYVA-) UNIV VANDERBILT.
XX Heilerqvist CG, Fu C;
XX WPI: 2000-205377/18.
DR P-PSDB: AA145088.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Claim 3; Page 83-86; 109pp; English.
XX
CC The present cDNA sequence encodes partial sheep GBS (group B beta
CC -haemolytic streptococci) toxin receptor (SP55). This sequence was cloned
CC using a primary culture of sheep lung endothelial cells. Expression
CC vectors comprising this cDNA can be transformed into host cells to
CC express GBS toxin receptor and its fragments. Detecting the receptor in
CC tissues is used to diagnose pathological vascularisation, e.g. for
CC detecting cancer metastases. GBS toxin receptors are useful for treating
CC neovascularisation (specifically cancer, reperfusion injury, scarring
CC during wound healing, keloids, chronic inflammation (rheumatoid
CC arthritis or psoriasis) or neural injury) and to raise specific
CC antibodies used for treating early onset disease. Inhibitors of this
CC receptor are useful for treating pathological or hypoxia-induced
CC endothelial cell proliferation and migration.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

Query Match 100.0%; Score 2844; DB 21; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgggggcggggggcttcggcgctccgctggagctctcttcccgcgagagctttg 60
Db 1 cccgggggcgggggcttcggcgctccgctggagctctcttcccgcgagagctttg 60
QY 61 cgcgcgtagctccctgaagagatcagagtcgcccggttcgacttagcccgagcgagc 120
Db 61 cgcgcgtagctccctgaagagatcagagtcgcccggttcgacttagcccgagcgagc 120
QY 121 gcgagggaggtcgcgcgcacacgcgtcctgcagcgcgcccgcgagcgagcgctc 180
Db 121 gcgagggaggtcgcgcgcacacgcgtcctgcagcgcgcccgcgagcgagcgctc 180
QY 181 cagatgctcgtcgtcgttacaactagatcttctgtcttcttgggttcttcgttc 240
Db 181 cagatgctcgtcgtcgttacaactagatcttctgtcttcttgggttcttcgttc 240
QY 241 tctatctaatcgggtgaatcgtgagcttgcactagtgagacatgtagattcaaacaa 300
Db 241 tctatctaatcgggtgaatcgtgagcttgcactagtgagacatgtagattcaaacaa 300
QY 301 ctgcacaaagataagacgtcctaagtggtgcagagcatctgtccataaagttc 360
Db 301 ctgcacaaagataagacgtcctaagtggtgcagagcatctgtccataaagttc 360
QY 361 ttccacaaacgaaggttaaaaagatcgcgtggatgcagaacactcaaggtatgtctcg 420
Db 361 ttccacaaacgaaggttaaaaagatcgcgtggatgcagaacactcaaggtatgtctcg 420
QY 421 gatcttttctatggtacatcacacaattctctggaagatatgtgccaagaa 480
Db 421 gatcttttctatggtacatcacacaattctctggaagatatgtgccaagaa 480
QY 481 gtgaggggagagctgtgtagattcgcggatcttgcataagctatcaccctgttca 540
Db 481 gtgaggggagagctgtgtagattcgcggatcttgcataagctatcaccctgttca 540
QY 541 ctcccccgtgcagattctcgaggtcgagccctgttgactcaaggcactagaagggc 600
Db 541 ctcccccgtgcagattctcgaggtcgagccctgttgactcaaggcactagaagggc 600

QY 601 taagagaggtgtcacatccagccatgcattgcattgttctcaatgggtcccccctc 660
Db 601 taagagaggtgtcacatccagccatgcattgcattgttctcaatgggtcccccctc 660
QY 661 ttgaagaagaagcctctcgtagatttcatactcagaagacacaacttggagcagttt 720
Db 661 ttgaagaagaagcctctcgtagatttcatactcagaagacacaacttggagcagttt 720
QY 721 ctctctctcttcgtgagtaatttgcactatgtatgtacttattgtcttatttct 780
Db 721 ctctctctcttcgtgagtaatttgcactatgtatgtacttattgtcttatttct 780
QY 781 ttgcatgttggaatcatctgttttaattatgagctccttagttagtgcacaccag 840
Db 781 ttgcatgttggaatcatctgttttaattatgagctccttagttagtgcacaccag 840
QY 841 aaactcaagaacaatcactccctatgaaaaggatattcttccatcataaaaaaanc 900
Db 841 aaactcaagaacaatcactccctatgaaaaggatattcttccatcataaaaaaanc 900
QY 901 agctctcttcaagaagtcagtcggtggttacctatgctgaatacactgcacatttgg 960
Db 901 agctctcttcaagaagtcagtcggtggttacctatgctgaatacactgcacatttgg 960
QY 961 ctattgtcgttgacatttcttcaacacgtgactttaaacttgttaccattttgc 1020
Db 961 ctattgtcgttgacatttcttcaacacgtgactttaaacttgttaccattttgc 1020
QY 1021 ctacttaacatgaagaagtcctaaagtlcaattcaagaagaatgggttttatctgcag 1080
Db 1021 ctacttaacatgaagaagtcctaaagtlcaattcaagaagaatgggttttatctgcag 1080
QY 1081 tcccttaattaggttctgtgttatgatatgctctgcgggtcaagctgcgaacaattta 1140
Db 1081 tcccttaattaggttctgtgttatgatatgctctgcgggtcaagctgcgaacaattta 1140
QY 1141 gggcaagatgaaatttctcaacctctgggtctgaaagagttttaaaccctatagagatga 1200
Db 1141 gggcaagatgaaatttctcaacctctgggtctgaaagagttttaaaccctatagagatga 1200
QY 1201 ttggaactgcgataatcccggttgcgcgagattatagcgttgatcttcccttgcgtc 1260
Db 1201 ttggaactgcgataatcccggttgcgcgagattatagcgttgatcttcccttgcgtc 1260
QY 1261 ttgatttccataacatcaacaacccctggagagcttctgcctctctgattagaata 1320
Db 1261 ttgatttccataacatcaacaacccctggagagcttctgcctctctgattagaata 1320
QY 1321 accatctggaatctgtcctctgtatgctgtgtaattcctcgtggacatcaacaataccttg 1380
Db 1321 accatctggaatctgtcctctgtatgctgtgtaattcctcgtggacatcaacaataccttg 1380
QY 1381 ccaactatccctggaatgattgggccatcaatttgcagaagctcttaccctggagaacta 1440
Db 1381 ccaactatccctggaatgattgggccatcaatttgcagaagctcttaccctggagaacta 1440
QY 1441 ttggaagaatggcaactgtttcttcgacatcgctgctatcaatgtatttggvccaatt 1500
Db 1441 ttggaagaatggcaactgtttcttcgacatcgctgctatcaatgtatttggvccaatt 1500
QY 1501 tcttcaacactatcgcgcaagtgaaagtgcacaacgtgggcatcagtgatccacaagagc 1560
Db 1501 tcttcaacactatcgcgcaagtgaaagtgcacaacgtgggcatcagtgatccacaagagc 1560
QY 1561 acagaacttgaaggaaaccaataataatccctgtctctatataatgtatcttcttcatc 1620
Db 1561 acagaacttgaaggaaaccaataataatccctgtctctatataatgtatcttcttcatc 1620
QY 1621 gtaactaaagtgccttggatattttaaagtgtgaagcaatctatataaagaataaat 1680
Db 1621 gtaactaaagtgccttggatattttaaagtgtgaagcaatctatataaagaataaat 1680

QY 1681 gtaataaagaaattgtttagatattgttaaggcttgtaatacgaatgaatgtcactagttgccc 1740
|||||
Db 1681 gtaataaagaaattgtttagatattgttgaaggcttgtaatacgaatgaatgtcactagttgccc 1740
QY 1741 atataagcaaaatagactatttttaattattataaaccggttgcgtgaacttaattcc 1800
|||||
Db 1741 atataagcaaaatagactatttttaattattataaaccggttgcgtgaacttaattcc 1800
QY 1801 aaggtaacatctgtgctgaacgttcaagccaaccacaagaaggagttattattat 1860
|||||
Db 1801 aaggtaacatctgtgctgaacgttcaagccaaccacaagaaggagttattattat 1860
QY 1861 aagaccataacctaaagaagaatgaagctgaataagacccttcataccttgccttaattaagg 1920
|||||
Db 1861 aagaccataacctaaagaagaatgaagctgaataagacccttcataccttgccttaattaagg 1920
QY 1921 tggataataattctcaagcttgcgttgaataacatctgttttgcataccttcctcaaaaatt 1980
|||||
Db 1921 tggataataattctcaagcttgcgttgaataacatctgttttgcataccttcctcaaaaatt 1980
QY 1981 attgtgatacgaacatccctggaacatgaagttcctcaactttagccttcacaggaagcttg 2040
|||||
Db 1981 attgtgatacgaacatccctggaacatgaagttcctcaactttagccttcacaggaagcttg 2040
QY 2041 cagccactgtatcatcactcaagccttggaacttcaacttggaaggaagcctgcccaggaagctgccc 2100
|||||
Db 2041 cagccactgtatcatcactcaagccttggaacttcaacttggaaggaagcctgcccaggaagctgccc 2100
QY 2101 acatgtccctctctgcgttcaaggaacagtcccaagcactttaggaagcactccaagaccag 2160
|||||
Db 2101 acatgtccctctctgcgttcaaggaacagtcccaagcactttaggaagcactccaagaccag 2160
QY 2161 ggtcagaagcgaagagctttaggaagcgtatcttccctcggggcgtttagttgtgtgaag 2220
|||||
Db 2161 ggtcagaagcgaagagctttaggaagcgtatcttccctcggggcgtttagttgtgtgaag 2220
QY 2221 cccctgagccaacagaggaagcagcgatcccaagatcagtttccatcacccctccctc 2280
|||||
Db 2221 cccctgagccaacagaggaagcagcgatcccaagatcagtttccatcacccctccctc 2280
QY 2281 ccttcccaagcaacatggaatattgctggaatgaactgcaaaagaagtgtgatgct 2340
|||||
Db 2281 ccttcccaagcaacatggaatattgctggaatgaactgcaaaagaagtgtgatgct 2340
QY 2341 aattgacacataataacatcatccttgatgatccttcaacttcaacttgaatgaatgaatgaat 2400
|||||
Db 2341 aattgacacataataacatcatccttgatgatccttcaacttcaacttgaatgaatgaatgaat 2400
QY 2401 tcttcaagtcctgtgttcttagagagctacacccaagaataataataaagaagaacaa 2460
|||||
Db 2401 tcttcaagtcctgtgttcttagagagctacacccaagaataataataaagaagaacaa 2460
QY 2461 acatttttctgtctgacacctaagtgtctgtgtcagttcaagttaaagttaacttcta 2520
|||||
Db 2461 acatttttctgtctgacacctaagtgtctgtgtcagttcaagttaaagttaacttcta 2520
QY 2521 ctccacaataacccggaacccgggtgttaatacatcctttagttgtgtttttaaattttat 2580
|||||
Db 2521 ctccacaataacccggaacccgggtgttaatacatcctttagttgtgtttttaaattttat 2580
QY 2581 ttttccagtttctcctggaacatctccacagttgttttgagcagctttagaggaacttatt 2640
|||||
Db 2581 ttttccagtttctcctggaacatctccacagttgttttgagcagctttagaggaacttatt 2640
QY 2641 gttgattagaactgtatgaatgtttcaactgtgtatgttcaacactaccgttaatttta 2700
|||||
Db 2641 gttgattagaactgtatgaatgtttcaactgtgtatgttcaacactaccgttaatttta 2700
QY 2701 actaaagcatttaagttaataatgaatgtatgatcattcgttaataatttttaacttcg 2760
|||||
Db 2701 actaaagcatttaagttaataatgaatgtatgatcattcgttaataatttttaacttcg 2760
QY 2761 taaatagcttaagtgtcatgtgtatatttctttaaacaatcaataaatatttttgg 2820

Db 2761 taaatagcttaagtgtcatgtgtatatttctttaaacaatcaataaatatttttgg 2820
QY 2821 aatgataaaaaaaaaaaaaaaaaa 2844
Db 2821 aatgataaaaaaaaaaaaaaaaaa 2844
RESULT 2
AAD10326
ID AAD10326 standard; DNA; 2844 BP.
XX
AC AAD10326;
XX
DT 16-OCT-2001 (first entry)
XX
DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) DNA.
XX
KW Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;
KW cytostatic; vulnery; antiatherosclerotic; osteopathic; vasotrophic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine; ds.
XX
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT CDS 84..1571
FT /tag="a
FT /product="Sheep GBS toxin receptor protein (SP55)"
XX
PN MO200156598-A2.
XX
PD 09-AUG-2001.
XX
XX 02-FEB-2001; 2001MO-US03662.
PF
XX 02-FEB-2000; 2000US-0179870.
PR
XX (UYVA-) UNIV VANDERBILT.
PA
XX
PI
XX
XX
DR MPI: 2001-488844/53.
DR P-PSDB: AAE06519.
XX
XX
PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment -
XX
XX
PS Disclosure: Page 47-50; 52pp; English.
XX
XX
CC The present sequence is a DNA encoding sheep group B beta-haemolytic
CC Streptococci (GBS) toxin receptor protein, SP55. The present invention
CC relates to a method for preventing or attenuating a patho-angiogenic
CC condition in a mammal which comprises administering to the mammal one
CC or more GBS toxin receptors or their immunogenic fragments to induce
CC or maintain an immune response to one of GBS toxin receptors. The
CC method is useful for preventing or ameliorating pathoangiogenic
CC conditions such as cancer, scarring during wound healing, gliosis
CC during repair of nerve injury, chronic wounds, keloids, reperfusion
CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
CC psoriasis in mammals. The proteins of the invention are also used
CC as vaccines.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

Query Match 100.0%; Score 2844; DB 22; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccggggcgggggtctcgccggtcccgctggagctctcttccgcggaaggtttg 60
Db 1 cccggggcgggggtctcgccggtcccgctggagctctcttccgcggaaggtttg 60
QY 61 cggcgtagctccctgaaggcatctgaagtcgccggttccgacttagcccgagcg 120
Db 61 cggcgtagctccctgaaggcatctgaagtcgccggttccgacttagcccgagcg 120
QY 121 gcgaggggggtctggaccgcaaccgctctcgacgcccgcggcggaaccgcctc 180
Db 121 gcgaggggggtctggaccgcaaccgctctcgacgcccgcggcggaaccgcctc 180
QY 181 cagatctgctcgtctcgcttaccacactagcattttgctcttttggttcttgctc 240
Db 181 cagatctgctcgtctcgcttaccacactagcattttgctcttttggttcttgctc 240
QY 241 tctatctaatcgggtgaatctgaagcgttgcactagtggacatggtgattcaacaa 300
Db 241 tctatctaatcgggtgaatctgaagcgttgcactagtggacatggtgattcaacaa 300
QY 301 ctgccaagaataagaaagctccctacagagtgtagagagcatctgcctcccaataaagttc 360
Db 301 ctgccaagaataagaaagctccctacagagtgtagagagcatctgcctcccaataaagttc 360
QY 361 ttccaaccacaacggttaaaagtaacggtgggtgtagcagaactcaaggatggaatttc 420
Db 361 ttccaaccacaacggttaaaagtaacggtgggtgtagcagaactcaaggatggaatttc 420
QY 421 gatctttttctatgctacatcatcaacaactctctggagatagtggtcgacagaa 480
Db 421 gatctttttctatgctacatcatcaacaactctctggagatagtggtcgacagaa 480
QY 481 gtggggggaagcgtgttctaagattcgagatcttgtctacagcatcttcaaccctgtca 540
Db 481 gtggggggaagcgtgttctaagattcgagatcttgtctacagcatcttcaaccctgtca 540
QY 541 ctccctcgtctgcatcttcggaagtcgagccctgtgtgcatcaggggacatgaagaaggc 600
Db 541 ctccctcgtctgcatcttcggaagtcgagccctgtgtgcatcaggggacatgaagaaggc 600
QY 601 taggaagaggtgtacacatcatccagcatgcatggtgtctctcatgggtccctcc 660
Db 601 taggaagaggtgtacacatcatccagcatgcatggtgtctctcatgggtccctcc 660
QY 661 ttgaaagaagcaagctctcgatattccatctgacagagcaacactgggagcagtagtt 720
Db 661 ttgaaagaagcaagctctcgatattccatctgacagagcaacactgggagcagtagtt 720
QY 721 ctctctctcttctggaagtaattgtctactatagaaatggaactatgctctctattct 780
Db 721 ctctctctcttctggaagtaattgtctactatagaaatggaactatgctctctattct 780
QY 781 ttggcatgtttggatatactggtttattttatgatactgcttagtagatagacacag 840
Db 781 ttggcatgtttggatatactggtttattttatgatactgcttagtagatagacacag 840
QY 841 aaactcaagaacatacctccctatgaaaaaggatatactcttcaatgaataaactc 900
Db 841 aaactcaagaacatacctccctatgaaaaaggatatactcttcaatgaataaactc 900
QY 901 agctctctcaagaagtcagtgccgtggatgacatgctgaactgcaactgcactttggg 960
Db 901 agctctctcaagaagtcagtgccgtggatgacatgctgaactgcaactgcactttggg 960
QY 961 ctattgtcgttgacaattttcttaacaactggacttttatacttattgacattatgc 1020
Db 961 ctattgtcgttgacaattttcttaacaactggacttttatacttattgacattatgc 1020
QY 1021 ctacttacaagaagagtccttaaggttcaatatcaagaagaatgggttttattctgag 1080
Db 1021 ctacttacaagaagagtccttaaggttcaatatcaagaagaatgggttttattctgag 1080

QY 1081 tcccttattaggttgtgtatgatagtacatctctcggtcgaaagctgacacatttaa 1140
Db 1081 tcccttattaggttgtgtatgatagtacatctctcggtcgaaagctgacacatttaa 1140
QY 1141 gggcaagaatggaattttcaactctgtggttgcgaaggttttagccttataaggatga 1200
Db 1141 gggcaagaatggaattttcaactctgtggttgcgaaggttttagccttataaggatga 1200
QY 1201 ttggaccttgatattctctgttgcgcaagatttaaggctggtatattctctgtgct 1260
Db 1201 ttggaccttgatattctctgttgcgcaagatttaaggctggtatattctctgtgct 1260
QY 1261 ttgcatctccaataccaacaacccctggagagctttctctctctggaatttagca 1320
Db 1261 ttgcatctccaataccaacaacccctggagagctttctctctctggaatttagca 1320
QY 1321 accatctggaacatgctctctgtaigtgtattctccctgggacatacaataactttg 1380
Db 1321 accatctggaacatgctctctgtaigtgtattctccctgggacatacaataactttg 1380
QY 1381 ccaactctctggaatgatttgggcccatacttgccagaagctcttaaccctggagaacata 1440
Db 1381 ccaactctctggaatgatttgggcccatacttgccagaagctcttaaccctggagaacata 1440
QY 1441 ttggaagaatggcaactgtttctgcatcgcgtgctgatacaatglatttggtgccatt 1500
Db 1441 ttggaagaatggcaactgtttctgcatcgcgtgctgatacaatglatttggtgccatt 1500
QY 1501 tcttcaacatcttcgccaagggtgaagtgcaaaaactgggccatgagatcaccaaggac 1560
Db 1501 tcttcaacatcttcgccaagggtgaagtgcaaaaactgggccatgagatcaccaaggac 1560
QY 1561 acaagaactgaaggaaccaataaatactcgtctctattaaagtatcttggatttaccat 1620
Db 1561 acaagaactgaaggaaccaataaatactcgtctctattaaagtatcttggatttaccat 1620
QY 1621 gtaacctaaagtgcttggatattttaatctgtgaagcaatcatatacaagaataaatt 1680
Db 1621 gtaacctaaagtgcttggatattttaatctgtgaagcaatcatatacaagaataaatt 1680
QY 1681 gtaactaaanaaatgtgtatgatttgaaggtgtgatacatgaagaatgtaactgtgctc 1740
Db 1681 gtaactaaanaaatgtgtatgatttgaaggtgtgatacatgaagaatgtaactgtgctc 1740
QY 1741 atataagcaaaatagcatattttaattatataacccgttggctggaacttaaaattc 1800
Db 1741 atataagcaaaatagcatattttaattatataacccgttggctggaacttaaaattc 1800
QY 1801 aggttcaacatactgctgcaagtcaggaacccaataagggaggttctattattat 1860
Db 1801 aggttcaacatactgctgcaagtcaggaacccaataagggaggttctattattat 1860
QY 1861 aagaacctaaactaaagaagatgagctgaaatagaaccttataaccttggtaattaag 1920
Db 1861 aagaacctaaactaaagaagatgagctgaaatagaaccttataaccttggtaattaag 1920
QY 1921 ttgataaattctcaagcttgtttaaacatctgttttttacaaccttcccaaaaaatt 1980
Db 1921 ttgataaattctcaagcttgtttaaacatctgttttttacaaccttcccaaaaaatt 1980
QY 1981 attgtlcatgaagcaatccctgacatgtaggttctcaaaactttagccttccacggagctg 2040
Db 1981 attgtlcatgaagcaatccctgacatgtaggttctcaaaactttagccttccacggagctg 2040
QY 2041 cagccactgtatacatcagccctggcaacttcaactgaggaagcatgcccagagcagctgc 2100
Db 2041 cagccactgtatacatcagccctggcaacttcaactgaggaagcatgcccagagcagctgc 2100
QY 2101 acatgtccctctctggtctcaggaagtgcccaagacacttaaggcaagcatcaagacag 2160
Db 2101 acatgtccctctctggtctcaggaagtgcccaagacacttaaggcaagcatcaagacag 2160
QY 2161 gtcagcgccaagccttggacgtaattcttcccttggggctgttaagtgtgtatgaag 2220

```
Db 2161 ggtcagcgccgaagcgttgcagcgtatcttcctccctggcggttaagtgctgagag 2220
Oy 2221 cccctgagcacaagagagagcgagtcacacagtcagtcagttccatgccccttccttc 2280
Db 2221 cccctgagcacaagagagagcgagtcacacagtcagtcagttccatgccccttccttc 2280
Oy 2281 ccttcacgacacgtgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 2340
Db 2281 ccttcacgacacgtgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 2340
Oy 2341 aattgacacatataacatcatccttgatgacacacacacacacacacacacacacacac 2400
Db 2341 aattgacacatataacatcatccttgatgacacacacacacacacacacacacacacac 2400
Oy 2401 tcttcaagtcctggttcttgaagtcacacacacacacacacacacacacacacacacac 2460
Db 2401 tcttcaagtcctggttcttgaagtcacacacacacacacacacacacacacacacacac 2460
Oy 2461 acatttctctgctgacacacacacacacacacacacacacacacacacacacacacac 2520
Db 2461 acatttctctgctgacacacacacacacacacacacacacacacacacacacacacac 2520
Oy 2521 cttcacacacacacacacacacacacacacacacacacacacacacacacacacacac 2580
Db 2521 cttcacacacacacacacacacacacacacacacacacacacacacacacacacacac 2580
Oy 2581 tttccacagtcctgctgacacacacacacacacacacacacacacacacacacacacac 2640
Db 2581 tttccacagtcctgctgacacacacacacacacacacacacacacacacacacacacac 2640
Oy 2641 gtgatttgaactgattgaattgattgattgattgattgattgattgattgattgattgatt 2700
Db 2641 gtgatttgaactgattgaattgattgattgattgattgattgattgattgattgattgatt 2700
Oy 2701 actaaagcatttaagttaagttaagttaagttaagttaagttaagttaagttaagttaagt 2760
Db 2701 actaaagcatttaagttaagttaagttaagttaagttaagttaagttaagttaagttaagt 2760
Oy 2761 taatagtcatttaagttaagttaagttaagttaagttaagttaagttaagttaagttaagt 2820
Db 2761 taatagtcatttaagttaagttaagttaagttaagttaagttaagttaagttaagttaagt 2820
Oy 2821 aatgataaaaaaaaaaaaaa 2844
Db 2821 aatgataaaaaaaaaaaaaa 2844
```

RESULT 3
AAF55900
ID AAF55900 standard; DNA; 2512 BP.

XX AC AAF55900;
XX DT 18-APR-2001 (first entry)
XX DE Human AST coding sequence.
XX KW Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
XX KW anion and sugar transporter; anion-cation symporter;
XX KW sialic acid transporter; ss.
XX OS Homo sapiens.
XX FN EPI069184-A1.
XX PD 17-JAN-2001.
XX PE 16-JUL-1999; 99EP-0202341.
XX PR 16-JUL-1999; 99EP-0202341.
XX PA (ALKU) AKZO NOBEL NV.

```
XX WP1: 2001-193090/20.  
DR P-PSDB; AAB66967.  
XX PT New human transporter gene implicated in Salla disease and lysosomal  
PT sialic acid transport, useful in assays for identifying new drugs, or  
PT diagnosing sialic acid transport defects related to mutations in the  
XX transporter gene  
XX PS Claim 2; Page 12-13; 20pp; English.  
XX CC The present sequence is the coding sequence for human Anion and Sugar  
CC Transporter (AST) protein. AST has significant homology with several  
CC members of the anion-cation symporter (ACS) family of transporters. AST  
CC is implicated in Salla disease, and is useful in screening assays for  
CC identifying new drugs. Compounds identified via AST screening is useful  
CC for preparing a pharmaceutical suitable as an activator or inhibitor of a  
CC sialic acid transporter protein. The pharmaceutical may be used in sialic  
CC acid associated diseases and CNS/immune related disorders.  
XX SO Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other.  
  
Query Match 55.3%; Score 1572.4; DB 22; Length 2512;  
Best Local Similarity 82.1%; Pred. No. 0;  
Matches 1917; Conservative 0; Mismatches 386; Indels 33; Gaps 8;  
  
Oy 5 ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 64  
Db 194 ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 253  
Oy 65 gtatgctccctgaagcagcatgaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 124  
Db 254 ct--gctacgtaagcgctcaagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 311  
Oy 125 ggaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 184  
Db 312 ggaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 371  
Oy 185 atgctgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 244  
Db 372 gtgcgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 431  
Oy 245 ttcatcaggggtgaatcgcagtcgctgctgctgctgctgctgctgctgctgctgctgctgctg 304  
Db 432 tgcattacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 491  
Oy 305 caaagataaagaaagcgtcctcagagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 364  
Db 492 agaagataaagaaagcgtcctcagagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 551  
Oy 365 caaccaaagcgtgtaaaagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 424  
Db 552 taatcaaagcgtgtaaaagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 611  
Oy 425 ttttctatgctacacacacacacacacacacacacacacacacacacacacacacacacacac 484  
Db 612 ctcttctatgctacacacacacacacacacacacacacacacacacacacacacacacacacac 671  
Oy 485 ggggaagtcgtgctagatcgcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgag 544  
Db 672 ggggaagtcgtgctagatcgcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgag 731  
Oy 545 cctgcgtcagatcgcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 604  
Db 732 catgctgcagatcgcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 791  
Oy 605 agaaggttcacacacacacacacacacacacacacacacacacacacacacacacacacacac 664  
Db 792 agaaggttcacacacacacacacacacacacacacacacacacacacacacacacacacacac 851  
Oy 665 aagaagcagctcgaattcctatcagcaggaacacacacacacacacacacacacacacacacac 724
```

Db 852 aagaagcaaacctcttagcatcttcataatgacagacagcttgggacgaattctct 911
 QY 725 tccctctctcgagtaattgctactataatgaaatcgactatgctcttcattcttgg 784
 Db 912 tccctctctcgagtaattgctactataatgaaatcgactatgctcttcattcttgg 971
 QY 785 catgttggatacatctggtttatttaatgacgtctgactatgtagtataaccagaac 844
 Db 972 taataatgagataattggttcttcttcttggatctgtagttagttagttagttagttagt 1031
 QY 845 tcaacagacatacctccctctctctctctctctctctctctctctctctctctct 904
 Db 1032 acaacagagaaattcccatatgaaagaaatacattctctctctctctctctctctct 1091
 QY 905 ctcttcacagaatgcagtcgctgtaatacctatgctgtaatacctgcaacttgggctat 964
 Db 1092 tcttcacagaatgcagtcgctgtaatacctatgctgtaatacctgcaacttgggctat 1151
 QY 965 tgcgttgcacatttcttacaactggaacttttaacttcttgaacttctgctac 1024
 Db 1152 cgtatgtgcacatttcttacaactggaacttttaacttcttgaacttctgctac 1211
 QY 1025 ttacatgaagaagtcctaagttcaatatcagaagaatgggtttttatctgcaatgcc 1084
 Db 1212 ttatagaagaagatcccaaggttcaatgctcagaagaatgggtttttatctctatggcc 1271
 QY 1085 ttattagatgtgtgtatgataatcctgctgctgctgctgctgctgctgctgctgctg 1144
 Db 1272 ttattagatgtgtgtatgataatcctgctgctgctgctgctgctgctgctgctgctg 1331
 QY 1145 aagaatgaaatttcaactctgctgctgctgctgctgctgctgctgctgctgctg 1204
 Db 1332 aagaatgaaatttcaactctgctgctgctgctgctgctgctgctgctgctgctg 1391
 QY 1205 acctgcgaatttccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1264
 Db 1392 acctgcgaatttccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1451
 QY 1265 attcctaaccatacacaacccctggaggtcttctctctctctctctctctctctct 1324
 Db 1452 ttctcctaaccatacacaacccctggaggtcttctctctctctctctctctctctct 1511
 QY 1325 tctggaacttgcctctctgctgctgctgctgctgctgctgctgctgctgctgctg 1384
 Db 1512 tctggaacttgcctctctgctgctgctgctgctgctgctgctgctgctgctgctg 1571
 QY 1385 tattcctggaatgattgggcccctctctctctctctctctctctctctctctctct 1444
 Db 1572 tattcctggaatgattgggcccctctctctctctctctctctctctctctctctct 1631
 QY 1445 agaatggcaaacctgcttctgctgctgctgctgctgctgctgctgctgctgctgct 1504
 Db 1632 agaatggcaaacctgcttctgctgctgctgctgctgctgctgctgctgctgctgct 1691
 QY 1505 cacactattgcgcaaaagtgaagtgaacacctggtgcaactgtagtaccacaaaggaac 1564
 Db 1692 tacaattatgcgcaaaagtgaagtgaacacctggtgcaactgtagtaccacaaaggaac 1751
 QY 1565 aaactgaagaagaacaataataatcctgctctctctctctctctctctctctctct 1624
 Db 1752 aaactgaagaagaacaataataatcctgctctctctctctctctctctctctctct 1811
 QY 1625 cctaaagaatgcttctgataatttaagtgtgaagcaatctcatatacaagaataaagtgtac 1684
 Db 1812 cctaaagaatgcttctgataatttaagtgtgaagcaatctcatatacaagaataaagtgtac 1857
 QY 1685 tagaagaatgtgttagattgtgaagctgtgaatcatgaatgtcatagttgtccatat 1744
 Db 1858 aatgtgactgttagattgtgaagctgtgaatcatgaatgtcatagttgtccatat 1917
 QY 1745 aagaagaatgtgctatttataatataatataccgctgtgctggaactataaactgaagg 1804
 Db 1918 taataaatagaactgtgttaataatataatataatgaatgaatgaatgaatgaatgaatga 1977

QY 1805 tcaacatactgctgtgaagtcaggaacccacacataagaggagttctcatlta-lltaag 1863
 Db 1978 tcaacataccgctgctgtgagtcgagcaacatgaatgagagagtcgtgtgatttttag 2037
 QY 1864 accatacctaagaagatgagctgaataagacccctctataccttcttctaataaagtgg 1923
 Db 2038 gcatactaaaggaatgagctgaataagacccctctgataccttcttctaataaactag 2097
 QY 1924 atataatctcaggtctgtttaaacaatcgttctgtgataccttctcacaanaatatt 1983
 Db 2098 atgataatctcaggtctgtttaaacaatcgttctgtgataccttctcacaanaatatt 2152
 QY 1984 tgcatacgaacatccctgacatgtagtctcaaacctttagcctctccacagagctggag 2043
 Db 2153 -atgtgagctctctctgtaacttagaccccaactttagcctctctgagagctgcat 2211
 QY 2044 ccactgtatcatcagctgtgcaacttcaactgaggaagcatgcccagagctgcca 2103
 Db 2212 ccactgtatcatcagctgtgcaacttcaactgaggaagcatgcccagagctgcca 2271
 QY 2104 tgcctcctctgtgcttgcagggagcagtgcccagcactt-----agcagacatcca 2156
 Db 2272 cactccctcctgtgcttgcagggatgagtgcccagcgttatcagaagagcagatccaag 2331
 QY 2157 ccagggatcagcgccaaagcttggacggtatcttccct--ggagctgttaatgtgtg 2214
 Db 2332 ccagggatcagcgctgagctctctgctgctgctgctgctgctgctgctgctgctgctg 2391
 QY 2215 atgaagccctgagccaagaggaagcagatcagatcatggttccatgacccct- 2273
 Db 2392 atgaagccctgagccaagaggaagcagatcagatcatggttccatgacccct- 2451
 QY 2274 tcccttcccttcccaacacacactgagatgctgctgctgctgctgctgctgctgctg 2329
 Db 2452 aacttcccttcccaacacacactgagatgctgctgctgctgctgctgctgctgctg 2507

RESULT 4
 AAZ50879
 ID AAZ50879 standard; cDNA; 2930 BP.
 XX
 AC AAZ50879;
 XX
 DT 31-MAY-2000 (first entry)
 XX
 DE Full length human GBS toxin receptor (HP59) cDNA.
 XX
 KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
 KW pathological vascularisation; cancer metastases; angiodermatitis;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 263..1873
 FT /tag= a
 FT /product= "Human GBS toxin receptor protein"
 PN MO200005375-A1.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-0516676.
 XX
 PR 22-JUL-1998; 98US-0093843.
 XX
 PA (UYVA-) UNTV VANDERBILT.
 XX
 PI Helleqvist CG, Fu C;

Db 847 caltgtgcagatttagaggtltygaccactcatlgtactcagagcactaggaagctag 906
QY 605 agaggggtgcacatataccgacatgcgatgcgtgtgtcttcatgggctccctcttga 664
Db 907 agaggggtgttacatcttccgacatgcatgcatgtgtgtcttcttggctccctcttga 966
QY 665 aagaagaagctctgtgatttccatatalgcaagagacaaacttggagacagtagttctct 724
Db 967 aagaagaacactctttagattctgtatgcagagacagcttggagacagtagttctct 1026
QY 725 tccctcttctgaggaatttgcctactatataatgtgactatgtctctattcttcttg 784
Db 1027 tccctcttctggaataattgtctactatataatgtgactatgtctctattcttcttg 1086
QY 785 catgttgcacatcgtgttatttatgtatcgtgtgactgtgactgtatgataccagaac 844
Db 1087 tactattggaatatttgtgtcttcttcttcttcttcttcttcttcttcttcttctt 1146
QY 845 tcacaagaacatcaactcccatlgaagaagagatatacttcttcatatlaanaaatcagct 904
Db 1147 acacaagaagaattcccatatgaaagaagaatacattcttcttcatatlaanaaatcagct 1206
QY 905 ccttcacagaagtcagtgccgtgtgatacctatgtctgaaatacactgcacatttggctat 964
Db 1207 tcttcacagaagtcagtgccgtgtgatacctatgtctgaaatacactgcacatttggctat 1266
QY 965 tgtctgtgcacatttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1024
Db 1267 cgtatgtgcacacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1326
QY 1025 ttacatgaagaagtcctlaaggttcaataltcaagaagaatgggttcttcttctgcagctcc 1084
Db 1327 ttatagaagaagtcctlaaggttcaataltcaagaagaatgggttcttcttctgcagctcc 1386
QY 1085 ttattttaggtgtgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 1144
Db 1387 ttattttaggtgtgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 1446
QY 1145 aagatggaatttctaaactcgtgtgtgtcgaagatttcttcttcttcttcttcttcttctt 1204
Db 1447 aagatggaatttctaaactcgtgtgtgtcgaagatttcttcttcttcttcttcttcttctt 1506
QY 1205 accctgcgatattcctgt 1264
Db 1507 accctgcgatattcctgt 1566
QY 1265 attcctaaccatataaacaaccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1324
Db 1567 ttctctaactatataaacaacaaccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1626
QY 1325 tctggaacatgtctcctcgtatgcgtgtatgtctcctgtgcacataaataccttgcac 1384
Db 1627 tctggaacatgtctcctcgtatgcgtgtatgtctcctgtgcacataaataccttgcac 1686
QY 1385 tattcctggaatgt 1444
Db 1687 tattcctggaatgt 1746
QY 1445 agaattggaacaaatgttctcgtatgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1504
Db 1747 agaattggaacaaatgttctcgtatgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1806
QY 1505 caccatattcgcacaaaggtgaagtgcaaaactgtgcacatcgaatgtatcacaagagacag 1564
Db 1807 tacactattcgcacaaaggtgaagtgcaaaactgtgcacatcgaatgtatcacaagagacag 1866
QY 1565 aaactggaaggaacaaataataatcctcgtctatataatgtatccttcttcttcttcttct 1624
Db 1867 aaactggaaggaacaaataataatcctcgtctatataatgtatccttcttcttcttcttct 1926
QY 1625 ccttaaatgtccttctgtatatttaattgtgtgaagcactatataaataaataatgtgtac 1684
Db 1927 ccttaaatgtccttctgtatatttaattgtgtgaagcactatataaataaataatgtgtac 1972

QY 1685 tagaaaatgtgttagatttgaaggtgtgtaacatgtaaatgtacatgtgtccat 1744
Db 1973 aattgtactgttattagattttaaaggcctataatcaatgaataatcagttgtccagaa 2032
QY 1745 aagcaaatatgactatltttaaattatataccggttctgtcgtgaactatcaatccag 1804
Db 2033 taataaataagcaatgtgtttaaattatgaataatgtgaagcagcttcttctttagt 2092
QY 1805 ttacatattcgtcgtcaagtgcagcaaccacaatagaggatcttattta- ttataag 1863
Db 2093 tcaatcactcgtcgttagtcgggcaacatgaagtagagcagttcgtgtattttag 2152
QY 1864 accataccataagagatgtgcgaataagacacttcaacttcaacttgcgttaataagtg 1923
Db 2153 gccataactaaaggaaatgtgcgaacaaacactcctcgtataacttgcgttaataaactag 2212
QY 1924 ataataattcctcaggtcgtttaaacaatcgttcttgtacacacttccctcaaaaattat 1983
Db 2213 atgataattcctcaggtacgtataaacaactgtgtgttcttcttcttcttcttcttcttct 2267
QY 1984 tgtcatcagcaatcctcgtacatgtatgtatcacaacttagcctctccacgagctgtcag 2043
Db 2268 -atgtcagctcctcgtacacacttagacccaactttagcatctcgtgtgagctgtccat 2326
QY 2044 ccactgtatcattcagcctgcgaacttcaatgtaggaagcagccacagcagctgtccaa 2103
Db 2327 ccactgtatattcgtcgtgcgaactgtaggaagcagccacagcagctgtccaa 2386
QY 2104 tglccctcctcgtcgtcagggacagtgcccgacact-----aggagatccaa 2156
Db 2387 cactccctcctcgtcgtcagggatcagagtgccagcgtttatcagagagagatccaa 2446
QY 2157 ccaggtcagcgcgaagcttgtgacggtattcttccct--ggggcgttataatgtgtg 2214
Db 2447 ccagagcagtgctgactcttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2506
QY 2215 atgaagccctgtgacacagggacgcgcgtatccacagatcagtggtttccatgcacctc- 2273
Db 2507 ataaagccctgtgtaggcaagacagtgatccacactgtagtgcgtgtgtgtgtgtgtgtgt 2566
QY 2274 tcccttcccttccacacactgtgatttgcctgtgcagatgtaacctgcgaaga 2329
Db 2567 aacttcccttccacacagagagataatgtgtgcagatgtaacctgcgaaga 2622

RESULT 6
AAZ50875
ID AAZ50875 standard; cDNA: 2602 BP.
XX
AC AAZ50875;
XX
DT 31-MAY-2000 (first entry)
XX
DE Partial human GBS toxin receptor (HP55) cDNA.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 58..1545
FT CDS /*tag= a
FT /product= "Partial human GBS toxin receptor"
XX
PN W0200005375-A1.
XX
PD 03-FEB-2000.

OY	1695	gtgttaagatttgtaaggctgtgtatcaatgaatgtaactagttgcataatgaagcaaat	1794
Db	1655	gtataaattttaaaggctataatcatbtaaatatactagttgcgaataataataatg	1714
OY	1755	agcattcttaatatattatgaacccgtttgtgcgaactcaatcaatgaagtcaatact	1814
Db	1715	aactggttttaatatatgaataataatgtaagctagaactcttaattagttcaatact	1774
OY	1815	ggtctgaagtcagagcaaccacataataggggagttcatlta--ttataagaataacta	1873
Db	1775	gctcgctagtcaggcaacatgaagtagagcaagttctgttatttttagggcactacta	1834
OY	1874	aagagtagagctgaataaagaccctctctataactttgctaataatgaagggtgaataatc	1933
Db	1835	agggaaatgaagctgtaaaaagaccctctgtataactttgtcttaataaactagaataatlc	1894
OY	1934	tcaagctctctttaaacaactgcgtttttgttaacaacttccataaaaattattgtacagc	1993
Db	1895	tcaagtaactatgaataacacactgtgtgttcttaacttccataaaa-----atgttaagc	1948
OY	1994	aatccctcgacaatgttagttctcaacttttagcctctccacgagagctggaagccatgtatc	2053
Db	1949	tctctcgtaacaactaagaccctcaaacctttagcatctctgtgggcctgcacatccactata	2008
OY	2054	atccagccttgcaaacacttcaactttgtgggaagaacatgagccagctgcacatgtccacct	2113
Db	2009	atttcgccttgcaaacctggaactgtagggagttgtgcccagagcagctgcacaagcactccc	2068
OY	2114	ctgagcttcagagcaagctgcccagaactt-----agcagagatccaaagcacaaggtcag	2166
Db	2069	ctgagcttcagaggtctcagaggtgcccagagttatcatagggagagcatccaaagccagagccag	2128
OY	2167	cgcgaagagccttggaaaggtatttttccct--gggagctgtaagtgtggtgaagaagccct	2224
Db	2129	tgttcgaacctctggcctggtgctcttccctccctgaggggctatcaatgltgataaagccct	2188
OY	2225	gagcgaacagaggaagcagcgatcaacagatcatggtttccatcacacctc--tccctccct	2283
Db	2189	gagtaggcaagagcagtgagatctccactgctatagttcttgatatacatccccaacttccct	2248
OY	2284	tcccaagcacactggaatgattgtcctgacatgtaacctgcataaagaana	2329
Db	2249	tcccaagacagagaatattggtctgtgcatagcacaactgcataaagaana	2294

```

1  RESULT      7
2  AAK94876
3  ID          AAK94876 standard; cDNA; 2712 BP.
4  XX
5  AC
6  AC          AAK94876;
7  XX
8  DT          06-NOV-2001 (first entry)
9  DT
10 DE          Human full-length cDNA, SEQ ID NO: 4068.
11 DE
12 XX          Human, full length cDNA; cDNA synthesis; oligo-capping; ss.
13 KM
14 XX
15 OS          Homo sapiens.
16 OS
17 XX          EP1130094-A2.
18 PN
19 XX
20 PD          05-SEP-2001.
21 PD
22 XX
23 PF          07-JUL-2000; 2000EP-0114089.
24 PF
25 XX          08-JUL-1999; 99JP-0194486.
26 XX          11-JAN-2000; 2000JP-0118774.
27 PR          02-MAY-2000; 2000JP-0183765.
28 PR
29 PA
30 (HELI-) HELIX RES INST.
31
32 Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

```

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
DR P-PSDB; AAM93914.
XX
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX
PS Claim 8; SEQ ID NO 4068; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 2712 BP; 677 A; 619 C; 619 G; 797 T; 0 other;
SQ

Query Match	51.3%	Score 1458.8	DB 22	Length 2712
Best Local Similarity	78.6%	Prod. No. 0	Mismatches 387	Indels 135
Matches 1916	Conservative	0	Gaps	9
OY	5	ggggggggggttcctggcggtcccgcttgagctcccttccgcggagcagatttggcc	64	
Db	6	ggccggggggtctggcggtctggcggtcccgcgctccctctctgcgaagtggcgattaac	65	
OY	65	gtatgccctgaaggcatcatgaaagttcccggttctggacttaagcccgagcgacgca	124	
Db	66	ct- g ctacgtgaaggcgatcatagttcgtccggttcgagacctgcccgaacgatgycga	123	
OY	125	ggagggtcggaacccgacacgcgtctctgcaagcgccccggcgggggaacccgctcaggt	184	
Db	124	ggagggtcggaacccgacacgcgtctctcctcctcagcgggccccacgggacggcgcctcaggt	183	
OY	185	atgcgtcctctgctctgcttaacactagaattttgtcctcttttggttcttcgtctcta	244	
Db	184	gtgcgtcctctgctctgcttaacactagaattttgtcctttttgtttcttcattgtta	243	
OY	245	ttcattaggggtggaatcctgagcgtttgcctagtggaccttggttgatttaaacacacgtgc	304	
Db	244	tgcatctaggtggaatcctgagcgtttgtcgtttagtgatattglaatttaacatacaacttc	303	
OY	305	caaaagataagaaagtcctcctcgaagtgagcagagatctgcctccataaaagtcttcca	364	
Db	304	agaagataatagaacttcccaagcggtgctccagagagattctgtccacataaaagtctacca	363	
OY	365	caaccaaaccgggttaaaaaagtaccgggtgtagatgcagaactcaagagatgtcttcgcgactc	424	
Db	364	taattcaaaacgggttaagaagttcccaattggatgagcagaaaaactcaagatgtattcttcggttc	423	
OY	425	ttttctcatatgctacatcatcacacaatctccctggaagatattgttcacagagaagtg	484	
Db	424	ctttttatattgctacatcatcacacaagaattccctggagagatattgttcgagcaaaatagg	483	
OY	485	ggggaagctgtgtcaggatctcggatctttgtctacagatccttcacacctgttcaactcc	544	
Db	484	gggggaaaatgtcgtcagagattttgagatccctgtgcactctgtccctcacccgttcaactcc	543	
OY	545	ctctgcctcagatctcggatccgagacccctgtttgcaattcagggcacagagaaggtcagg	604	
Db	544	catgtctcgagatttaggaagtttggaacccactatgttactcagagacatagaagaactagg	603	
OY	605	agaagggtgcacataccagccatgacatgacatgtgtctcatatggtctcccccctctga	664	
Db	604	agaagggttcatattccagccatgacatgacatgtgtgtctctttaggtctccccccttga	663	

QY 665 aagaagcaagcttcgtatctcatalatgcagagcaaaccttggagcaagtctctct 724
D 664 aagaagcaaaccttcgtatctcatalatgcagagcaaaccttggagcaagtctctct 723
QY 725 tccctctctgagtaattgtctactatagaaattgtaactatgctctctctctg 784
D 724 tccctctctgagtaattgtctactatagaaattgtaactatgctctctctctg 783
QY 785 catgttggaatcatctggtatttattatgagctgctggttggtatgatacagaagac 844
D 784 taactatggaatcatctggtatttattatgagctgctggttggtatgatacagaagac 843
QY 845 tcacaagaacaatcatctcctactatgaaagagatattcttcacatcaataaagac 899
D 844 aacaagaagaattcccatatagaaagaaatatactcttcacatcaataaagac 903
QY 900 ----- 899
D 904 tgaagtgcagtagatgatcttggtctcagctccctctgctcccggttcaagc 963
QY 900 ----- 922
D 964 catctcccaactcagctcctcaagtagttggaagcaagcttctctcaagaagtagt 1023
QY 923 gccgtggaactactatgctgaatcactgcacttgggtatgtgctggtgacatttctc 982
D 1024 gccgtggaactactatgctgaatcactgcacttgggtatgtgctggtgacatttctc 1083
QY 983 ttaacaactgagcttctatattctgtgacatttgcctactatcaatgaagaagctct 1042
D 1084 ttaacaactgagcttctatattctgtgacatttgcctactatcaatgaagaagctct 1143
QY 1043 aaggttcaaatcaagaagaatggtgttttattctgcagctccctctcttattagttgtt 1102
D 1144 aaggttcaaatcaagaagaatggtgttttattctgcagctccctctcttattagttgtt 1203
QY 1103 atgataatccctgcctgcaggtcgaagctgcgaacttgaaggaagcaagatgttctaac 1162
D 1204 atgataatccctgcctgcaggtcgaagctgcgaacttgaaggaagcaagatgttctaac 1263
QY 1163 tctgtggttctcgaagaatttttagccttataaggaatgttgaacctgcagatctctgt 1222
D 1264 tctgtggttctcgaagaatttttagccttataaggaatgttgaacctgcagatctctgt 1323
QY 1223 tgcgcgcaggaattatagagctggtgtattctcctggtgcattccttaacatataaac 1282
D 1324 agctgcgcaggaattatagagctggtgtattctcctggtgcattccttaacatataaac 1383
QY 1283 aacctggaagcttctgtctctcttgattatgacataacacatctgacattgctctc 1342
D 1384 aacctggaagcttctgtctctcttgattatgacataacacatctgacattgctctc 1443
QY 1343 gtagtgcgtgattctccttggtgcatacaaatattctgtcacattctctggaatgttg 1402
D 1444 gtagtgcgtgattctccttggtgcatacaaatattctgtcacattctctggaatgttg 1503
QY 1403 gcccatcatgccaagaagcttcaacctggaagaactattggaagaatggaatgttct 1462
D 1504 gcccatcatgccaagaagcttcaacctggaagaactattggaagaatggaatgttct 1563
QY 1463 ctgcatactgctgctgctatcaatgatttgggtgcatttcttcaacatctgcgcaag 1522
D 1564 ctgcatactgctgctgctatcaatgatttgggtgcatttcttcaacatctgcgcaag 1623
QY 1523 tgaagtgcacaactggtgcacatgtagtcaacaaggaagcaagaactggaaggaacata 1582
D 1624 tgaagtgcacaactggtgcacatgtagtcaacaaggaagcaagaactggaaggaacata 1683
QY 1583 aataatcctctctctatataatgtagtattctgttataatgtaactaaagtgccttgat 1642
D 1684 aataatcctctctctatataatgtagtattctgttataatgtaactaaagtgcctt--- 1739

QY 1643 atttaattggtgaagcaatcatatatacaagaataaattgtactagaataattgtgtaga 1702
D 1740 -ctgtattggtgaagcaatcatatatacaagaataaattgtactagaataattgtgtaga 1789
QY 1703 ttgtgaagctgtgataatcaatgaatgtcactagttgcctataatgaagaattagctatt 1762
D 1790 ttgtgaagctgtgataatcaatgaatgtcactagttgcctataatgaagaattagctatt 1849
QY 1763 ttaatttatttaacccgttgcgtggaactataactcaaggtgcataactgctgctga 1822
D 1850 ttaatttatttaacccgttgcgtggaactataactcaaggtgcataactgctgctga 1909
QY 1823 gtcaggaacccaataagggagttcattta-ttataagaacatacctaagaagatg 1881
D 1910 gtcaggaacccaataagggagttcattta-ttataagaacatacctaagaagatg 1969
QY 1882 agctggaatgagccctctactacttgccttaattgaatggtgataaataatctcaggtc 1941
D 1970 agctggaatgagccctctactacttgccttaattgaatggtgataaataatctcaggtc 2029
QY 1942 tgttaaacatctgtttgtgaacactcctcccaaaaattattgttcaatgaacatccctg 2001
D 2030 tgttaaacatctgtttgtgaacactcctcccaaaaattattgttcaatgaacatccctg 2083
QY 2002 acatgtagtctcaaatattagcctctcacaaggagctgcagcaactgtatcatcagcc 2061
D 2084 acatgtagtctcaaatattagcctctcacaaggagctgcagcaactgtatcatcagcc 2143
QY 2062 tgcgaactcaactgaggaagcaatgcccagcgctgcacatgctccctctgctc 2121
D 2144 tgcgaactcaactgaggaagcaatgcccagcgctgcacatgctccctctgctc 2203
QY 2122 agggacagctgcagcaact-----agggacacccaagcaaggtgcagcccaag 2174
D 2204 agggacagctgcagcaact-----agggacacccaagcaaggtgcagcccaag 2263
QY 2175 ctgtgagcgttatctctccct--ggggcgttaattgtgtga-tgaagccctgaagcaac 2232
D 2264 ctgtgagcgttatctctccct--ggggcgttaattgtgtga-tgaagccctgaagcaac 2323
QY 2233 agggacagcgcgatcccaatcaatggtttccatgacacccctc-ccctccctcccaag 2291
D 2324 agggacagcgcgatcccaatcaatggtttccatgacacccctc-ccctccctcccaag 2383
QY 2292 cactgagatattgcgtgcagtgtaacctgcgaaga 2329
D 2384 cagaggaatattgctgctgcacactgcgaaga 2421

RESULT 8
AAZ50880
ID AAZ50880 standard; cDNA; 1485 BP.
XX
AC AAZ50880;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human/sheep consensus GBS toxin receptor cDNA sequence-1.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritic; anti-psoriatic; ss.
OS Homo sapiens.
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..1485
FT /tag= a
FT /product= "Human/sheep consensus GBS toxin receptor

protein sequence"
FT /transl_except- (pos:4..6; aa:Xaa)
FT /transl_except- (pos:7..9; aa:Xaa)
FT /transl_except- (pos:16..18; aa:Xaa)
FT /transl_except- (pos:22..24; aa:Xaa)
FT /transl_except- (pos:28..30; aa:Xaa)
FT /transl_except- (pos:31..33; aa:Xaa)
FT /transl_except- (pos:34..36; aa:Xaa)
FT /transl_except- (pos:46..48; aa:Xaa)
FT /transl_except- (pos:49..51; aa:Xaa)
FT /transl_except- (pos:58..60; aa:Xaa)
FT /transl_except- (pos:61..63; aa:Xaa)
FT /transl_except- (pos:64..66; aa:Xaa)
FT /transl_except- (pos:67..69; aa:Xaa)
FT /transl_except- (pos:70..72; aa:Xaa)
FT /transl_except- (pos:73..75; aa:Xaa)
FT /transl_except- (pos:79..81; aa:Xaa)
FT /transl_except- (pos:85..87; aa:Xaa)
FT /transl_except- (pos:91..93; aa:Xaa)
FT /transl_except- (pos:100..102; aa:Xaa)
FT /transl_except- (pos:124..126; aa:Xaa)
FT /transl_except- (pos:130..132; aa:Xaa)
FT /transl_except- (pos:136..138; aa:Xaa)
FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..198; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..252; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)
FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)

FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)
FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)
FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1081..1083; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)

DE	Human/Sheep consensus GBS toxin receptor cDNA sequence-2.	
XX		
KW	Human GBS toxin receptor; group B beta-haemolytic streptococci; Hp59;	
KW	pathological vascularisation; cancer metastases; angiogenesis;	
KW	neovascularisation; reperfusion injury; scarring; keloid;	
KW	chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;	
KW	endothelial cell proliferation; antibacterial; anticancer;	
KW	anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss	
XX		
OS	Homo sapiens.	
OS	Ovis sp.	
XX		
EH	Key	Location/Qualifiers
FT	mat_peptide	1..1485
FT		/*tag= a
FT		/product= "Human/sheep consensus GBS toxin receptor
FT		protein sequence"
FT		/transl_except= (pos:4..6; aa:Xaa)
FT		/transl_except= (pos:7..9; aa:Xaa)
FT		/transl_except= (pos:16..18; aa:Xaa)
FT		/transl_except= (pos:22..24; aa:Xaa)
FT		/transl_except= (pos:28..30; aa:Xaa)
FT		/transl_except= (pos:31..33; aa:Xaa)
FT		/transl_except= (pos:34..36; aa:Xaa)
FT		/transl_except= (pos:46..48; aa:Xaa)
FT		/transl_except= (pos:49..51; aa:Xaa)
FT		/transl_except= (pos:58..60; aa:Xaa)
FT		/transl_except= (pos:61..63; aa:Xaa)
FT		/transl_except= (pos:64..66; aa:Xaa)
FT		/transl_except= (pos:67..69; aa:Xaa)
FT		/transl_except= (pos:70..72; aa:Xaa)
FT		/transl_except= (pos:73..75; aa:Xaa)
FT		/transl_except= (pos:79..81; aa:Xaa)
FT		/transl_except= (pos:85..87; aa:Xaa)
FT		/transl_except= (pos:91..93; aa:Xaa)
FT		/transl_except= (pos:100..102; aa:Xaa)
FT		/transl_except= (pos:124..126; aa:Xaa)
FT		/transl_except= (pos:130..132; aa:Xaa)
FT		/transl_except= (pos:136..138; aa:Xaa)
FT		/transl_except= (pos:154..156; aa:Xaa)
FT		/transl_except= (pos:157..159; aa:Xaa)
FT		/transl_except= (pos:163..165; aa:Xaa)
FT		/transl_except= (pos:169..171; aa:Xaa)
FT		/transl_except= (pos:181..183; aa:Xaa)
FT		/transl_except= (pos:187..189; aa:Xaa)
FT		/transl_except= (pos:190..192; aa:Xaa)
FT		/transl_except= (pos:196..198; aa:Xaa)
FT		/transl_except= (pos:202..204; aa:Xaa)
FT		/transl_except= (pos:211..213; aa:Xaa)
FT		/transl_except= (pos:220..222; aa:Xaa)
FT		/transl_except= (pos:223..225; aa:Xaa)
FT		/transl_except= (pos:235..237; aa:Xaa)
FT		/transl_except= (pos:241..243; aa:Xaa)
FT		/transl_except= (pos:244..246; aa:Xaa)
FT		/transl_except= (pos:250..252; aa:Xaa)
FT		/transl_except= (pos:277..279; aa:Xaa)
FT		/transl_except= (pos:280..282; aa:Xaa)
FT		/transl_except= (pos:283..285; aa:Xaa)
FT		/transl_except= (pos:295..297; aa:Xaa)
FT		/transl_except= (pos:304..306; aa:Xaa)
FT		/transl_except= (pos:337..339; aa:Xaa)
FT		/transl_except= (pos:340..342; aa:Xaa)
FT		/transl_except= (pos:346..348; aa:Xaa)
FT		/transl_except= (pos:367..369; aa:Xaa)
FT		/transl_except= (pos:394..396; aa:Xaa)
FT		/transl_except= (pos:397..399; aa:Xaa)
FT		/transl_except= (pos:406..408; aa:Xaa)
FT		/transl_except= (pos:409..411; aa:Xaa)
FT		/transl_except= (pos:412..414; aa:Xaa)
FT		/transl_except= (pos:430..432; aa:Xaa)
FT		/transl_except= (pos:433..435; aa:Xaa)
FT		/transl_except= (pos:436..438; aa:Xaa)

FT	/trans_except-	(pos:442..444; aa:Xaa)
FT	/trans_except-	(pos:445..447; aa:Xaa)
FT	/trans_except-	(pos:463..465; aa:Xaa)
FT	/trans_except-	(pos:475..477; aa:Xaa)
FT	/trans_except-	(pos:481..483; aa:Xaa)
FT	/trans_except-	(pos:487..489; aa:Xaa)
FT	/trans_except-	(pos:490..492; aa:Xaa)
FT	/trans_except-	(pos:493..495; aa:Xaa)
FT	/trans_except-	(pos:496..498; aa:Xaa)
FT	/trans_except-	(pos:502..504; aa:Xaa)
FT	/trans_except-	(pos:514..516; aa:Xaa)
FT	/trans_except-	(pos:529..531; aa:Xaa)
FT	/trans_except-	(pos:535..537; aa:Xaa)
FT	/trans_except-	(pos:562..564; aa:Xaa)
FT	/trans_except-	(pos:589..591; aa:Xaa)
FT	/trans_except-	(pos:595..597; aa:Xaa)
FT	/trans_except-	(pos:598..600; aa:Xaa)
FT	/trans_except-	(pos:604..606; aa:Xaa)
FT	/trans_except-	(pos:619..621; aa:Xaa)
FT	/trans_except-	(pos:614..616; aa:Xaa)
FT	/trans_except-	(pos:651..653; aa:Xaa)
FT	/trans_except-	(pos:691..693; aa:Xaa)
FT	/trans_except-	(pos:694..696; aa:Xaa)
FT	/trans_except-	(pos:700..702; aa:Xaa)
FT	/trans_except-	(pos:703..705; aa:Xaa)
FT	/trans_except-	(pos:706..708; aa:Xaa)
FT	/trans_except-	(pos:712..714; aa:Xaa)
FT	/trans_except-	(pos:715..717; aa:Xaa)
FT	/trans_except-	(pos:724..726; aa:Xaa)
FT	/trans_except-	(pos:727..729; aa:Xaa)
FT	/trans_except-	(pos:736..738; aa:Xaa)
FT	/trans_except-	(pos:748..750; aa:Xaa)
FT	/trans_except-	(pos:757..759; aa:Xaa)
FT	/trans_except-	(pos:760..762; aa:Xaa)
FT	/trans_except-	(pos:769..771; aa:Xaa)
FT	/trans_except-	(pos:772..774; aa:Xaa)
FT	/trans_except-	(pos:775..777; aa:Xaa)
FT	/trans_except-	(pos:778..780; aa:Xaa)
FT	/trans_except-	(pos:790..792; aa:Xaa)
FT	/trans_except-	(pos:793..795; aa:Xaa)
FT	/trans_except-	(pos:811..813; aa:Xaa)
FT	/trans_except-	(pos:820..822; aa:Xaa)
FT	/trans_except-	(pos:847..849; aa:Xaa)
FT	/trans_except-	(pos:850..852; aa:Xaa)
FT	/trans_except-	(pos:853..855; aa:Xaa)
FT	/trans_except-	(pos:856..858; aa:Xaa)
FT	/trans_except-	(pos:862..864; aa:Xaa)
FT	/trans_except-	(pos:880..882; aa:Xaa)
FT	/trans_except-	(pos:883..885; aa:Xaa)
FT	/trans_except-	(pos:892..894; aa:Xaa)
FT	/trans_except-	(pos:922..924; aa:Xaa)
FT	/trans_except-	(pos:928..930; aa:Xaa)
FT	/trans_except-	(pos:943..945; aa:Xaa)
FT	/trans_except-	(pos:952..954; aa:Xaa)
FT	/trans_except-	(pos:955..957; aa:Xaa)
FT	/trans_except-	(pos:970..972; aa:Xaa)
FT	/trans_except-	(pos:994..996; aa:Xaa)
FT	/trans_except-	(pos:997..999; aa:Xaa)
FT	/trans_except-	(pos:1009..1011; aa:Xaa)
FT	/trans_except-	(pos:1012..1014; aa:Xaa)
FT	/trans_except-	(pos:1033..1035; aa:Xaa)
FT	/trans_except-	(pos:1063..1065; aa:Xaa)
FT	/trans_except-	(pos:1081..1083; aa:Xaa)
FT	/trans_except-	(pos:1084..1086; aa:Xaa)
FT	/trans_except-	(pos:1090..1092; aa:Xaa)
FT	/trans_except-	(pos:1096..1098; aa:Xaa)
FT	/trans_except-	(pos:1111..1113; aa:Xaa)
FT	/trans_except-	(pos:1126..1128; aa:Xaa)
FT	/trans_except-	(pos:1129..1131; aa:Xaa)
FT	/trans_except-	(pos:1138..1140; aa:Xaa)
FT	/trans_except-	(pos:1141..1143; aa:Xaa)
FT	/trans_except-	(pos:1144..1146; aa:Xaa)
FT	/trans_except-	(pos:1147..1149; aa:Xaa)

FT	/transl_except=	(pos:1150..1152; aa:Xaa)
FT	/transl_except=	(pos:1153..1155; aa:Xaa)
FT	/transl_except=	(pos:1158..1170; aa:Xaa)
FT	/transl_except=	(pos:1174..1176; aa:Xaa)
FT	/transl_except=	(pos:1180..1182; aa:Xaa)
FT	/transl_except=	(pos:1189..1191; aa:Xaa)
FT	/transl_except=	(pos:1201..1203; aa:Xaa)
FT	/transl_except=	(pos:1246..1248; aa:Xaa)
FT	/transl_except=	(pos:1270..1272; aa:Xaa)
FT	/transl_except=	(pos:1291..1293; aa:Xaa)
FT	/transl_except=	(pos:1306..1308; aa:Xaa)
FT	/transl_except=	(pos:1315..1317; aa:Xaa)
FT	/transl_except=	(pos:1324..1326; aa:Xaa)
FT	/transl_except=	(pos:1330..1332; aa:Xaa)
FT	/transl_except=	(pos:1333..1335; aa:Xaa)
FT	/transl_except=	(pos:1339..1341; aa:Xaa)
FT	/transl_except=	(pos:1348..1350; aa:Xaa)
FT	/transl_except=	(pos:1357..1359; aa:Xaa)
FT	/transl_except=	(pos:1372..1374; aa:Xaa)
FT	/transl_except=	(pos:1375..1377; aa:Xaa)
FT	/transl_except=	(pos:1381..1383; aa:Xaa)
FT	/transl_except=	(pos:1384..1386; aa:Xaa)
FT	/transl_except=	(pos:1396..1398; aa:Xaa)
FT	/transl_except=	(pos:1402..1404; aa:Xaa)
FT	/transl_except=	(pos:1420..1422; aa:Xaa)
FT	/transl_except=	(pos:1444..1446; aa:Xaa)
FT	/transl_except=	(pos:1456..1458; aa:Xaa)
FT	/transl_except=	(pos:1459..1461; aa:Xaa)
FT	/transl_except=	(pos:1462..1464; aa:Xaa)
FT	/transl_except=	(pos:1471..1473; aa:Xaa)
FT	/transl_except=	(pos:1483..1485; aa:Xaa)
FT	/note="Xaa = unknown"	

PN	WO200005375-A1.			
XX				
ED	03-FEB-2000.			
XX				
PF	22-JUL-1999;	99WO-US16676.		
XX				
PR	22-JUL-1998;	98US-0093843.		
XX				
PA	(UYVA-) UNIV VANDERBILT.			
XX				
PI	Hellerqvist CG, Fu C;			
XX				
DR	WPI; 2000-205377/18.			
DR	P-PSDB; AAY45090.			
XX				
Query Match	45.9%;	Score 1305;	DB 21;	Length 1485;
Best Local Similarity	87.9%;	Pred. No. 5e-293;		
Matches 1305; Conservative	0;	Mismatches 180;	Indels 0;	Gaps

QY	84	atgaagtcgccggttcctcggaattagcccgagcagcgaggaaggctcgaccgcaca	143
Db	1	atgagtcnccggtttmngacatnctgcconganccgagcgagsggncngaacgcacn	60
QY	144	ccgctctctgaagcgcgcccgcgagcggaacccgctccagtatgctctctgtctgtac	203
Db	61	ccnccnccnccnngcgcccccgcggcngaaacccgctccaglnctgctgcgcgttac	120
QY	204	aacctgaacattctgtccctttctgtgttctctgtctctcatcattacgagtgaaatcg	263
Db	121	aactnagcgaantctgcacctttctgtctctcttntntnatnccattagngtgaatcg	180
QY	264	agcgctgcacacagcagcagatgctgattccaacacacatgcacaaagtaatagaagctc	323
Db	181	agngttcgmntaagcagaaatgctgtatctccaanaacactmnaagataatagaacttcc	240
QY	324	tacgaagtgcagcagcattctgtctccataaaagtcttccacaaccacacgggtlaanaag	383
Db	241	nanngtgtctmagaagcattctgcgtcccatataaagtcttcanaancacacacgggtlaanaag	300

QY	384	taccggtggagtcagaaactcaaggatctcgatcttcctcgatcttcttctatctgctacatc	443
DB	301	taccmthtggatctgcagaactcaagatggatctctcgcnlctttttnatgtgtacatc	360
QY	444	atcaacaatctctctggaggtatggttcgacgaagaagtggggggaagctgtgtcaggaa	503
DB	361	atcaacaatctctctggaggtatggttcgacgaagaagaagtggggggaagctgtgtcaggaa	420
QY	504	ttcgagatctctgtcagatcatctctcaaccgtttcaactccccctcgttcagatcttcga	563
DB	421	ttnggagatccttgmaccngctcctntcaccctgttcaactccctnngctcgtcagattngga	480
QY	564	gtcgagacccttgttcacctaagggacatagaagggtagagaaggtgtcacatatcca	623
DB	481	gtngagancctnnttgnactcagtcagactagaagcctagaagagggtgtinatctcca	540
QY	624	ggcatcgatcgcatgtggtctctaatggcctccccctcttgaagaagcagactctcggt	683
DB	541	ggcagcgatcgcatgtggtctctaatggcctccccctcttgaagaagcagactctcna	600
QY	684	attcaatgtcaggaagcaacaacttggagacgtatgtctctctctctcttcgttgagtaat	743
DB	601	attcctntatcgagaagcaacacttggagacgtatgtctctctctctcttcgttgagtaat	660
QY	744	tgtactataltgaatttgacttatgtctctactctcttggcatgtgttgaaactctgg	803
DB	661	tgtactataltgaatttgacttatgtctctactctcttggnaanttttggaaatntnbtg	720
QY	804	tttatattatgatactcgtcttaagtatgtataccacgaagaactcacagaacatcacctcc	863
DB	721	ttntttttnbtgactctgntttagttagtgcacacacanaamncacagaanaatntnncm	780
QY	864	tatgaaagaggtatctcttcttcaataaaatccagctctcttcacagaagcagtg	923
DB	781	tatgaaagaggtatctcttcttcaataaaatccagctctcttcacagaagcagtg	840
QY	924	ccgtgagatccatgactgaaatcaactgcaccttgggcatactgtcgttgcacatttct	983
DB	841	ccgtgagatccatgaaatcaactgcaccttgggcatactgtcgttgcacatttct	900
QY	984	tacaactggacttttatacttgtgtgacttatcttgccactacatgaaaggaatccta	1043
DB	901	tacaactggacttttatacttnttgaacttatcttgcactatnataatgaaaggaatccta	960
QY	1044	aggttcaatatcgaagaatgggttttatactgcagcagctccattatgaagtgtgtgta	1103
DB	961	aggttcaatatcgaagaatgggttttatactgcagcagctccattatgaagtgtgtgta	1022
QY	1104	tgtatgatactgtctgggttcaagctgctgcacaatttaagggcaagatggaattttcaact	1163
DB	1021	tgtatgatactgtctgggttcaagctgctgcacaatttaagggcaagatggaattttcaact	1080
QY	1164	ctgtgggtctcgaaagatttttaagcctatataggatgattggacctgcgatalctctggt	1222
DB	1081	ntnbtgnttgcgaagantttttagcctatagatgagatgtggacctgcgmatlctctgtn	1140
QY	1224	ggcgaagattatagagctgtgattatctctgcgttgcattccataaccatacaca	1283
DB	1141	ggcgaagattatagagctgtgattatctctgcgttgcattccataaccatacaca	1200
QY	1284	accctgggaagcttctgtctctctcgatttagcatcaacatctgacatctgtctctctg	1343
DB	1201	acncttggagagcttctgtctctctcgatttagcatcaacatctgacatctgtctctctg	1260
QY	1344	tatgctgttatctctcctgggacatcaaaatccttgcacatatctccgtggaatgtgg	1403
DB	1261	tatgctgttatctctcctgggacatcaaaatccttgcacatatctccgtggaatgtgg	1320
QY	1404	cccatcatctcgaaagattcttaccctctggaagaacatctggagaatgtgaaacttttc	1465
DB	1321	cccatcatctcgaaagattcttaccctctggaagaacatcttggagaatgtgaaacngnttct	1380
QY	1464	tgcatacctctgtatcaaatgtatgttgggtgcacatttctcaacatctgcgcaaaagt	1521

Db 961 aggtcaatgttcaagaagaatgggttttattccttcattgtccttatttagtcttgytla 1020
QY 1104 tgaatgacccgtcgcggcgaagctgactgacatttaaggcgaagtgaattttcaact 1163
Db 1021 tgaatgacccgtcgcggcgaagctgactgacatttaaggcgaagtgaattttcaact 1080
QY 1164 ctgtgtgtcgaaggttttttagccttaagagatgttgacccgtgatatctctgtt 1223
Db 1081 ttatgtgtcgcagaatttttagccttaagagatgttgacccgtgatatctctgtt 1140
QY 1224 gccgagagattatagctgtgattattcctctgtctgttgcattcctcaatataca 1283
Db 1141 gctgtgtcctcatgtgtgtgattattcctctgtctgttgcattcctcaatataca 1200
QY 1284 accctggagagctttgtctctctgtgatttagcattcaaccctctggacattgtccctcg 1343
Db 1201 acactggagagctttgtctctctgtgatttagcattcaaccctctggacattgtccctcg 1260
QY 1344 taagctgtattctctctgtgacatacaaatccttgcacattcctctgaatgttgg 1403
Db 1261 tatgtgtatccctcctcctgacatacaaatccttgcacattcctctgaatgttgg 1320
QY 1404 cccatcatcgcgaagaagcttaaccctctgagaacactatgtgaatggcaactgttttc 1463
Db 1321 cccgtcatcgttaaaagctcgaaccctctgatacaactgttggagaatggcaaacgtgttc 1380
QY 1464 tgcctcgtctgctcatcaatgtatttggctgacatttcttcacacattcgcgcaaaagt 1523
Db 1381 tatattgtctgtctatataagtttttggctgacatttcttcacacattcgcgcaaaagt 1440
QY 1524 gaagtgcgaagaactggccatcagtgatcaccaagaagacagaactga 1571
Db 1441 gaagtgcgaagaactggccatcagtgatcaccaagaagacagaactga 1488

RESULT 11
AAH79234
ID AAH79234 standard; cDNA; 2670 BP.
XX
AC AAH79234;
XX
DT 20-NOV-2001 (first entry)
XX
DE Human sodium dependent phosphate co-transfer protein 35 cDNA.
XX
KW Human; sodium dependent phosphate co-transfer protein 35;
KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN CN1298882-A.
XX
PD 13-JUN-2001.
XX
PF 06-DEC-1999; 99CN-0124217.
XX
PR 06-DEC-1999; 99CN-0124217.
XX
PA (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WP1; 2001-503367/56.
DR P-PSDB; AAG65238.
XX
PS
XX
PT Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX
CC Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
CC
CC The present invention provides the protein and coding sequences of human
CC sodium dependent phosphate co-transfer protein 35. The sequences can be
CC used in the treatment of hypophosphaturia, hypercalcaemia,

CC hypophosphataemic rickets and nephritis. The present sequence is the
CC coding sequence of the invention.
XX
SQ Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;

Query Match 41.1%; Score 1169.8; DB 22; Length 2670;
Best Local Similarity 82.6%; Pred. No. 1.5e-261;
Matches 1434; Conservative 0; Mismatches 272; Indels 31; Gaps 7;

QY 604 gagaggggtgtaacatataccagcagcatgcatgctgctgcttctcattgctc 663
Db 1 gagaggggtgtaacatataccagcagcatgcatgctgctgcttctcattgctc 60
QY 664 aaaaagaagcttcagatattcctatgaggaagacaaacttggagagtagttctc 723
Db 61 aaaaagaagcttcagatattcctatgaggaagacaaacttggagagtagttctc 120
QY 724 ttctctctctgagtaatttgctactatataagatggaactatgctctctattcttg 783
Db 121 ttctctctctgagtaatttgctactatataagatggaactatgctctctattcttg 180
QY 784 gcaattgtggaatcactcgtgttattttagagatcgtcctagtagtgatacaccagaa 843
Db 181 gtaactatggaatatttctgttcttcttggatcgtgtagtagtgatacaccagaa 240
QY 844 ctcaagaagacatcccccataagaagaatattcttctcaatcaataaaatacagc 903
Db 241 aacccaagaagaatttccatatagaagaatcatalcttctcctcaataagaatacagc 300
QY 904 tctctcagaagaatcagtcgctgagatcactatgctgaatcactcgtccacttgggcta 963
Db 301 ttctcagaagaatcagtcgctgagatcactatgctgaatcactcgtccacttgggcta 360
QY 964 ttgtcgttgcacatttcttcaacactgagcttttattacttgttgaacttattgcta 1023
Db 361 tctgtatgacacttcttctcaacactgagcttttattacttgttgaacttattgcta 420
QY 1024 cttaacgaagaagctccaaagctcaatatacaagaagaatgggttttctctcagctc 1083
Db 421 cttaacgaagaagctccaaagctcaatatacaagaagaatgggttttctctcagctc 480
QY 1084 cttaattaggttctgttctatgatagtatcctgtcgtggaatcagctcgtgacaattaaagg 1143
Db 481 cttaattaggtcctctgttctatgatagtatcctgtcgtggaatcagctcgtgacaattaaagg 540
QY 1144 caaagtgaatttttcaacctgtggttcgaagaagtttttagcctataagagatgtg 1203
Db 541 caaagtgaatttttcaacctgtggttcgaagaagtttttagcctataagagatgtg 600
QY 1204 gacctgcgatatccctgtgtgcgcaagattttagcgtgttattctctgtgctgtg 1263
Db 601 gacctgcgatatccctgtgtgcgcaagattttagcgtgttattctctgtgctgtg 660
QY 1264 catcctaacaataatacaacaacctgggaagcttttgcctctctctgataatagatacaacc 1323
Db 661 catcctaacaataatacaacaacctgggaagcttttgcctctctctgataatagatacaacc 720
QY 1324 atctggacattgtctctctgtatgctgtgtattctctcgtggacatacaaatccttgcca 1383
Db 721 atctggacattgtctctctgtatgctgtgtattctctcgtggacatacaaatccttgcca 780
QY 1384 ctattcttggaatgattgggcccatactgacgaagaagcttaccgcccggaagaacatgct 1443
Db 781 ctattcttggaatgattgggcccatactgacgaagaagcttaccgcccggaagaacatgct 840
QY 1444 gagaatggcaaacgttctgtcatcgtctgctgctatcaatgatatgtgtgccaatttctc 1503
Db 841 gagaatggcaaacgttctgtcatcgtctgctgctatcaatgatatgtgtgccaatttctc 900
QY 1504 tcaacataatcgcgaagaagtggaagtgcaaaactggtgacatcagtgatcaccaaggacaca 1563
Db 901 tcaacataatcgcgaagaagtggaagtgcaaaactggtgacatcagtgatcaccaaggacaca 960


```
Db 1795 TCCTGGGCATCACAAATACATTTGCGCATCTATCCAGAAATGGTGGCCCGCATTCGTCTA 1736
Qy 1417 gaagcttacccttggaagaacatttggaagaacatgcttcttgatgcgcgtg 1476
Db 1735 AAGGTCTGACCCCTGATTAACACTGTTGGAGAAAGGCAACCGTGTCTATATTGCTGCTG 1676
Qy 1477 ctatcaatgatttggctgacatttcttcaca--ctatctgcacaagtgaaagtgcaaa 1534
Db 1675 CTATTAATGTTTTTGGGCAATTTCTTTAACTATTTCGCAAAAGGTGAATACAAA 1616
Qy 1535 ctgggcatcagtgatcaccagaagaacactgaaggaccacaataataatcctgtc 1594
Db 1615 CTGGGCTCTCAATGATCACCATGACACACACTGAAAGAACAAATAATATCTCTGCC 1556
Qy 1595 tctattaatgatatcttgattatcatgatactaaagtccttgatattttaaagtgt 1654
Db 1555 TCTATTAAATATTTTATTATTATCATGTAACCTCAAAAGTCCCTT-----CTGATTGTGT 1501
Qy 1655 aagcaatctatatacaagaataatgtaactagaaaaaattggttagatttgaagctt 1714
Db 1500 AAGCATTTCTATGT-----CTTTTTTAATGTACTGTATGATTTTAAGGCTT 1450
Qy 1715 gtaataatgaatgtaactgaattgcaatataagaacaaattagctattttaaattatt 1774
Db 1449 ATATATATGAATAATATCACTAGTTGCCAGAAATTAATAATGAACCTGTTTAATTAATGAAT 1390
Qy 1775 aacccttgctggaactcaatcgaaggtcacatatcgtgtggtgaagaagcaagcc 1834
Db 1389 AATATGTATAGCTAGGACTTCTTACTTAAAGTTTCATACATCTGCTGCTAGTGGGCAACAT 1330
Qy 1835 acaatagggaggtctatatta-cttataagaacatacctaagaagatgaagctgaataga 1893
Db 1329 GAAGTATAGACAGATTTCTGTTGATTTTATGAGGCGCATTAAGGAATGAGCTGAACACAGA 1270
Qy 1894 cccctctataccttcttgcttaataagtgatataatcttcaggtcctgttataacatct 1953
Db 1269 CCGCTGATACCTTTCTTAACTAATTAACATGATATATTCCTAGGTAACATTAACACCT 1210
Qy 1954 gttttgtacaccttccctcaaaaattattgtcatcagcaatccctgacatgtagctc 2013
Db 1209 GTTGTGTTTCATCTTCTCTCTATAAA-----ATTGTACAGCTCTCTGTACACTTAACCT 1156
Qy 2014 caaacttgaagctctccacagagctgycagcagctglatcatcagcctgycacacttcac 2073
Db 1155 CAACCTTTAGCATCTCTGTGGAGCTGCATCCACTGTATTAATTTCCCTGCGCAACTGGAC 1096
Qy 2074 tggaggaaagcatgcccagcagctgcccacatgtccctctctgcttccaggaagcagtgcc 2133
Db 1095 TGAGGGGAAGTGTGCCAGGAGCTGCAAGCACTCCCTCCCTGCTTCAGGGTCAAGAGTG 1036
Qy 2134 caagacctt-----aggcagcatccaagaacagggctcagcgcgaagccttggacgta 2186
Db 1035 CCAAGGCTTTATATAGAGGAGCATCCAGACCCAGAGCAGTGCAGCTCTTCGCTGGTG 976
Qy 2187 tctctccct-9gggcgtgtaattggtgataagccctgagcccaacagggagcagcgcg 2244
Db 975 CCTTCTCTGAGGGGCTATCATGTGTAGATAAGCCCTGAGTAGGCAAGACAGAGTGAG 916
Qy 2245 attcaagtcatggttccatgacccctc-tcccttcccttcccaagacaacatgagatatt 2303
Db 915 ATCCACTGCTATGCTGTGTATACATCTCAAACTTCCCTCCACAGACAGAGGAATATT 856
Qy 2304 gctgcatgtaacctgcacaaga 2329
Db 855 GCGTGGCATGCAACTGTCACAAAGAA 830
```

```
RESULT 13
AA159901
ID AA159901 standard: cDNA: 929 BP.
XX
AC AA159901;
```

```
XX 22-OCT-2001 (first entry)
DT
XX Human polynucleotide SEQ ID NO 3890.
DE
XX
XX Human: nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000MO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR
XX P-PSDB; AAM40745.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 3890; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytosarctic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 929 BP; 214 A; 224 C; 223 G; 268 T; 0 other;
```

```
Query Match 22.6%; Score 642.4; DB 22; Length 929;
Best Local Similarity 83.4%; Pred. No. 2.8e-139;
Matches 742; Conservative 0; Mismatches 146; Indels 2; Gaps 1;
```

```
Qy 5 ggagcgaggagcttcggcgctcccgctgagctctcttccgagagcaggttgcgcc 64
Db 42 ggcgcgggggcttcggcgcttcggcgcttccttcccttccgcaagtgagagtaacc 101
Qy 65 gtatgcctctgaagcatatgaagtcccggttccgacttagcccgagcgagcgaga 124
```


Db 484 ggggaaatgtctgtagattgggagatccttgctgctgctgctcctacccctgtcactcc 543
 Qy 545 cctgcgtgcagattcggagtcgagcccttgctgcactcagggcactagaagggtcagg 604
 Db 544 catgctgcagatttagagattgagccactcattgtactcagagcactagaaggactagg 603
 Qy 605 agagggtgacacatccagccatgcatgcatgtgcttcacatggtctccccccttga 664
 Db 604 agagggtgacacatccagccatgcatgcatgtgcttcacatggtctccccccttga 663
 Qy 665 aagaagaaagcttctgagtttctatgacagagacacacttggagcagatgtctct 724
 Db 664 aagaagaaagcttctgagtttctatgacagagacacacttggagcagatgtctct 723
 Qy 725 tccctctctgagaaattgtctactatgaaattgactatgtctcttctcttcttg 784
 Db 724 tccctctctgagaaattgtctactatgaaattgactatgtctcttctcttcttg 783
 Qy 785 catggttggaaatcactggttatttattgagatcgtcttagttagtgatatacacaagaac 844
 Db 784 nactatggaaaattgtggtcttcttngatctngtag-tagtgacacacacaacaaa 842
 Qy 845 tcacaagacaa 855
 Db 843 aaccagagaa 853
 RESULT 15
 AAK92364
 ID AAK92364 standard; cDNA; 838 BP.
 AC AAK92364;
 XX 06-NOV-2001 (first entry)
 DT
 DE Human cDNA 5'-end sequence, SEQ ID NO: 824.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 OS Homo sapiens.
 PN EP1130094-A2.
 PD 05-SEP-2001.
 PF 07-JUL-2000; 2000EP-0114089.
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI WPI; 2001-524255/58.
 DR
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS Claim 2; SEQ ID NO 824; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 5'-end of a cDNA provided in the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 838 BP; 189 A; 196 C; 210 G; 237 T; 6 other;

Query Match 19.4%; Score 553; DB 22; Length 838;
 Best Local Similarity 81.9%; Pred. No. 1.5e-118;
 Matches 660; Conservative 0; Mismatches 143; Indels 3; Gaps 2;

Qy 5 ggggaggggggttcggcgtcccgctgagctctcttccgggagcaggtttggcc 64
 Db 6 gggccgggggttcggcgtcccgctgagctctcttccgggagcaggtttggcc 65
 Qy 65 gtagctccctgaaagcactaagctcccggttctggaactaaccccgagcgagcgga 124
 Db 66 ct-gctcaagtagggcgtacatagatgtctccggttcgagacctggccggagcagtagcgga 123
 Qy 125 ggaagggtcggaacgcgacacgcgtctcagcgcgcccgcgaggcggaacccgtccagt 184
 Db 124 ggaagagcaggaacgcgacacgcgtctcagcgcgcccgcgaggcggaacccgtccagt 183
 Qy 185 atgctgctctgctcgttacaacactagcattttgctcttttggttctctgcttcta 244
 Db 184 gtgctgctctgctcgttacaacactagcattttgctcttttggttctctcattctgta 243
 Qy 245 ttcatcgggtgagatctgagcgttgcactagttgacatgtgacatgtacaaacacacatgc 304
 Db 244 tgcattacggtgagatctgagcgttgcactagttgacatgtgacatgtacaaacacatgc 303
 Qy 305 caaagataatagaacgtccctacagagtgtagaagacatctgctccataaagtcttca 364
 Db 304 agaagataatagaacgtccctacagagtgtagaagacatctgctccataaagtcttca 363
 Qy 365 caaccaaacgggttaaaagtagccggttgtagtgcagaacccaagatgagttctcgatc 424
 Db 364 taatcaaaacgggttaaaagtagccggttgtagtgcagaacccaagatgagttctcgatc 423
 Qy 425 tttttctatggtcactacatcaacacaaattcctctgagagatatgttccagcagaagtgg 484
 Db 424 cttttttatggtcactacatcaacacaaattcctctgagagatatgttccagcagaagtgg 483
 Qy 485 ggggaaagctgttgcagagattcgggacttctgtcacagctatcttcaacctgttcaacc 544
 Db 484 ggggaaagctgttgcagagattcgggacttctgtcacagctatcttcaacctgttcaacc 543
 Qy 545 cctgcgtgcagatttcggaatcggagccctgtgtgcactcagggcactagaagggtcagg 604
 Db 544 catgctgcagatttcggaatcggagccctgtgtgcactcagggcactagaagggtcagg 603
 Qy 605 aagaaggtgacacatccagccatgcatgcatgtgcttcacatggtctccccccttga 664
 Db 604 aagaaggtgacacatccagccatgcatgcatgtgcttcacatggtctccccccttga 663
 Qy 665 aagaagaaagcttctgagtttctatgacagagacacacttggagcagatgtctct 724
 Db 664 aagaagaaagcttctgagtttctatgacagagacacacttggagcagatgtctct 723
 Qy 725 tccctctctgagaaattgtctactatgaaattgactatgtctcttctcttcttg 784
 Db 724 tccctctctgagaaattgtctactatgaaattgactatgtctcttctcttcttg 782
 Qy 785 catggttggaaatcactggttattt 810
 Db 783 gactatggaatatttggttcttctt 808

Search completed: July 15, 2002, 13:53:18
 Job time: 14938 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 11:45:11 ; Search time 4323.72 Seconds
(without alignments)
8877.851 Million cell updates/sec

Title: US-09-776-865-3

Perfect score: 2844
Sequence: 1 cccgggggggggggggcttcg.....ataaaaaaaaaaaaaaaaaa 2844

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: qb_estl:*
10: qb_estl2:*
11: qb_hic:*
12: qb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	725.2	25.5	985	9	AL550137
2	508.6	17.9	734	10	BI907284
3	505.2	17.8	626	10	BF676817
4	504.4	17.7	801	10	BI661062
5	476	16.7	559	10	BM106101
6	469.6	16.5	787	10	BG198416
7	468.6	16.5	700	10	BE869819
8	460.2	16.2	736	10	BI860521
9	451.6	15.9	518	9	AM655673
10	448.4	15.8	711	10	BI697765
11	433.6	15.2	729	10	BE971208
12	426.6	15.0	477	10	BM106103
13	416	14.6	904	10	BF124137
14	413.8	14.5	676	10	BE867611
15	409.2	14.4	633	10	BI817031
16	407.4	14.3	895	10	BF539146
17	399	14.0	459	10	BF652406

18	395.8	13.9	682	9	BB613552
19	380.6	13.4	518	9	AA833297
20	375.8	13.2	689	10	BG400588
21	374.6	13.2	434	10	BF652410
22	368.8	13.0	645	9	BB537525
23	365.4	12.8	432	9	AA258513
24	359	12.6	630	9	BB610013
25	346	12.2	629	9	BB664731
26	335.2	11.8	965	10	BG290613
27	324	11.4	661	10	BG400668
28	322.8	11.4	1100	10	BG866603
29	317.6	11.2	360	10	BF042802
30	310.6	10.9	499	9	AL597124
31	304	10.7	320	10	BG687881
32	302	10.6	429	10	BF563945
33	300.4	10.6	658	10	BI851890
34	273.6	9.6	587	10	BM179717
35	272.4	9.6	721	10	BJ060500
36	270.2	9.5	524	10	BM272092
37	261.4	9.2	367	10	H63685
38	261	9.2	301	10	N31254
39	255.8	9.0	893	10	BG541099
40	254.6	9.0	632	10	BJ031727
41	246	8.6	360	10	BE663434
42	237.6	8.4	445	10	BE375421
43	233.8	8.2	376	9	AM486714
44	224.4	7.9	259	10	BI536090
45	219.6	7.7	321	9	AM531337

ALIGNMENTS

RESULT 1
AL550137 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.PL2 Homo sapiens cDNA clone CSOD104YK17 5
prime, mRNA sequence.
ACCESSION AL550137
VERSION AL550137.1 GI:12886813
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source 1..985

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSOD104YK17"

/clone_lib="LTI_NFL006.PL2"

/issue_type="Placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 236 a 224 c 300 t 1 others

ORIGIN

QY 366 aaccaaagggtaaaaagctaccggtggtgagtcagaaactcaagagtgattctcgatct 425
DB 359 AATCAAGCGGTAAAGATGATCCATGGATGGATGAGAACTCAAGATGATGATCTCGGTTCC 418
QY 426 ttctctatggtacatcatcatcacaatctcctggaagatagtgtccagcagaagtggtg 485
DB 419 TTTTATGAGGCTACATCATCTACACAGATCTCGAGAGGATGATTGTTGCCACGCAAAATAGGG 478
QY 486 gggaaagctgtctgagagtcggatctgtgtacagacttcaacctgttcaacctcc 545
DB 479 GGGAAATGCTCTAGCGATTTGGGATCCCTGGCACTGCTGCTCTACCTCTGTTCACTCC 538
QY 546 ctgcgtcgaagatctcggaagtcggaccctgtgtcactcaaggcactagaagtgtaga 605
DB 539 ATTGTGCGAGATTTAGGATTTGAGACCACTGATTTACTACAGACAGCTAAGAGACTAGGA 598
QY 606 gaggtgtcacatcatccatgcgatgcatgtgtcttctatgtgtccctcccttgaa 665
DB 599 GAGGATGTTACATTTCCACCCATGCAATGCAATGCTCTCTTGGGCTCCCTCTTGAA 658
QY 666 aaaaagaagctctgagatcttcatcatgcaagcacaacttggagacagt-aggttctct 724
DB 659 AGAACAACAACTCTTAGCATTTTCATATGACAGACACAACTTGGACAGTAATTTCTCT 718
QY 725 tctctcttctggaagtaattgtctactatagaattg 761
DB 719 TCCTCTTCTGGAATAA-TTGGTACTATATGAATTTGG 754

RESULT 3
BF676817 626 bp mRNA linear EST 21-DEC-2000
LOCUS 602084380F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:424881 5',
DEFINITION Bf676817
ACCESSION Bf676817
VERSION Bf676817.1 GI:11950712
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1067 row: 0 column: 20
High quality sequence stop: 598.
Location/Qualifiers
1. 626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:424881"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgctcgagc); Site_2: SfiI (ggcgatagcgc
1); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGCGCGGACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

FEATURES
Source

BASE COUNT 157 a 126 c 123 g 220 t
ORIGIN

Query Match 17.8%; Score 505.2; DB 10; Length 626;
Best Local Similarity 89.0%; Pred. No. 2.1e-82;
Matches 557; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

QY 1001 taacttggttaccattatctgacctacttacttgaaggaagtcctcaaggttcaatattcaaga 1060
DB 1 TACTTATTGACATTATTGCTTACTTATATGAGAGATCCCTAAGGTTCAATGTTCAAGA 60
QY 1061 gaatggttltatctgcgaagtccttatttaagtgtgtgtatglatatglatctgtcgg 1120
DB 61 GAATGGGTTTTATCTTATCTTATTCCTATTATAGCTCTTGATATGATGATCTGCTGCG 120
QY 1121 taaagctgtgcaattttaaaggaagatgtgaatttcaactctgtgtgttgcgaaggt 1180
DB 121 TCAACCTGCTGCAATTTAAAGGCAAAATGAAATTTTCACTTATATGCTGCGCAAT 180
QY 1181 tttagccttaaggatgattggaactggaatcttccgtgtgtgtgcgaagattatagg 1240
DB 181 TTTTAGCCTTATAGAAATGATTGACCTGCACTATTTCTGAGCTGCTGCTTATTTGG 240
QY 1241 ctgtatattcctctgtgtgtgtgtcatctcctaaccatataaaccctggagagctttg 1300
DB 241 CTTGATATTATCTTTGGCGGTGCTTCTCTAATATCAACACACTGGAGGCTTTTG 300
QY 1301 ctctctgattttagcatcaacaacatctggaatgtccctctgtatgtctgtatctcct 1360
DB 301 CTTCTGGAATTTACATCAACACATCTGGATTTCTCTCTGATGCTGATCTCTCT 360
QY 1361 gggcatcaaaatctcttccatcttccatcttccatcttccatcttccatcttccatcttcc 1420
DB 361 GGGCATCAAAATACATTTGGCCATTTCCATTTCCAGAAATGTTGGCCGCTGATTTGCTAAAG 420
QY 1421 tcttacccttggaagacactatgtggaatgtggaactgttctctgcatcgtcgtcat 1480
DB 421 TCTGACCCCTGATTAACACGCTTGGAGAAATGGCAACCGTCTTATTTGCTGCTGCTAT 480
QY 1481 caatgtatttgggtgcatttcttcaacatcttgcgaagtggaatgtggaactgtggtg 1540
DB 481 TAACTTTTGTGTCATTTCTTTTACACTATTTCCCAAGGTGAAGTAAACAAACTGGGC 540
QY 1541 catcgtatcatcccaagacagaagaactgtgaaggaacaaat-aaataatctgttctcat 1599
DB 541 TCTCATATATCCATGACACAGACACTGAAGGAACCAATTAATATCTGCTCTAT 600
QY 1600 taatgatcttctgttattcatgtaac 1625
DB 601 TTACTGATTTTATTATCATCTGTAAC 626

RESULT 4
B1661062 801 bp mRNA linear EST 12-SEP-2001
LOCUS 603304362F1 NIH_MGC_Mam4 Mus musculus cDNA clone IMAGE:5350046 5',
DEFINITION B1661062
ACCESSION B1661062
VERSION B1661062.1 GI:15575298
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 801)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.

Db 1 TAAATATTTCAGGCTTCTGTTAAACATCTGTTTTGTACCTTCCTCAAAAAA-TATTT 59
Qy 1985 gtcacagcaatccctgcacatgtagtctcaacttagcctccacaggaactgagcagc 2044
Db 60 GTCATCAGCAATCCCTGACACATAGTCTCAACTTTAGCCTCTTGAGCTAGCTGGCAGC 119
Qy 2045 cactatcatcactcagctcagcactcactcaggaagacatgcccagagcagctgcacat 2104
Db 120 CACTGTATCATTCAGCCCGCAACTTCACTGAGGAGAGCATGCCAGGAGAGCTGCCACAC 179
Qy 2105 gtccctctctgtgtcctcaggaagcagtgccagcacttaggcagcatccaagcagagtc 2164
Db 180 GTCCCTCTCTGTGGCTTACAGGAGACAGTGGCCACACTTAGGCGCATTCACAAGACCGGGTC 239
Qy 2165 agcgcgaagagcttgagcagctatcttccctcggagcgtgtaatbtytgatgaagccct 2224
Db 240 AGTGGCAAGGCTTTGGATGTAATTTCTTCCCTGGGGCTGTATATGTATGATGAAGCCCT 299
Qy 2225 gagcaacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2284
Db 300 GAGCCACAGGAGCAACTGATCCAGATCATGATTTCCATGCTACCCCTCCCT 359
Qy 2285 cccagcagcagctgagatgtcctcagcagcagcagcagcagcagcagcagcagcagcagc 2344
Db 360 CCCAGC-----GCATGTATCTGCAAAAAGAGTGTGATGCTTAAT 401
Qy 2345 agccacataacatcatcctctgtagcctcactcactcactcactcactcactcactcactc 2404
Db 402 AGCCACATATATCATCATCTTGTATGATCTTACCTTCACATGATGATCAAGTATTAATCTT 461
Qy 2405 caagctcgtgtctcaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2463
Db 462 CAAGCTGTGTGTAGAGAGTACACAGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 521
Qy 2464 ttttctgtctgacactaagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2501
Db 522 TTTTCTGTCTTACACCTTAAGTATCTGTGTGCACTTC 559

RESULT 6
Bg198416 787 bp mRNA 1linear EST 21-APR-2001
LOCUS Bg198416
DEFINITION RST17674 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION Bg198416
VERSION Bg198416.1 GI:13720103
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 787)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,
'E., Veloso,N., Klink,A., Hess,J., Cothen,K., Lo,K., Offenbacher,
'J., Danzig,J., and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 549.
Location/Qualifiers
1..787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

FEATURES
SOURCE

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 214 a 166 c 154 g 253 t
ORIGIN

Query Match 16.5%, Score 469.6; DB 10; Length 787;
Best Local Similarity 78.5%; Pred. No. 6.5e-76;
Matches 620; Conservative 0; Mismatches 149; Indels 21; Gaps 4;

Qy 1343 gtatgtgtatctccctcgggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1402
Db 3 GATCTGTGTATCTCTCGGGGATCACAATAATCATTTCCACTATTTCCAGGAATGGTTGG 62
Qy 1403 gccacatctgcagaagctctaccctcagagacacacttgagagatgcaactgtttc 1462
Db 63 GCCCGTCAATGTCTAATAAGCTGACCCCGATACACTGTGGAGATGGCAACCGTGT 122
Qy 1463 ctgcatcgtctgtctatcaatgatttgggtgcaatttcttcacactatcgcagaag 1522
Db 123 CCATATGCTGTCTATTAATGTTTGGCCCATTTTCTTACACATATTCGCCAAGG 182
Qy 1523 tgaagtgaagaactggccatcagtgatcacagaagcagcagcagcagcagcagcagcagc 1582
Db 183 TGAAGTACAAACGTGGCTCTCAATGATCACCATGACACAGCATGAGAACCAATA 242
Qy 1583 aataatctgtctcattatgatacttcttcatcagtaaacctaaagtgcccttgat 1642
Db 243 AATATTCCTGCTCATTAATGATATTTTATTTATCATGTAACCTCAAGTGCTT----- 298
Qy 1643 atttaatgtgaagaactctatatacaagalaatgtactagaataatgtgttaga 1702
Db 299 -CTGTATGTGTGAACCATATATATGT-----CTTTTAAATGTGATGTATATAGA 348
Qy 1703 ttgtgaagctgtgatatcagtaataatgactagtgccatlaagcaaatagattt 1762
Db 349 TTTTAAAGGCTATATATATGATAATATCATAGTGTGCGAATATTAATTAAGAACTGTGT 408
Qy 1763 ttaattatataacccgttctgtagaactacaactcagagtgccatcatctgtgtgca 1822
Db 409 TTAATTTAGAAATATATGTAAGCTAGACTCTACTTTAGGTGCATACACTGCTGCTA 468
Qy 1823 gtcaaggcaaccacaatagggagtgctattta-ttataagacacatccataagagatg 1881
Db 469 GTGGGGCAACATGAAGTGAAGCAGTTCTGTGATTTTAAAGGCCATCTAAAGGAATG 528
Qy 1882 agctgaatagacccctctatcactcttgcttaattgaagtgatgaataatctcagagct 1941
Db 529 AGCTGAACACAGACCTCCGATACCTTTGCTTAATTAACCTGATGATATTTCTCGAGTAC 588
Qy 1942 tgttaaacatctgttttctacaccttctcaaaaataattattgtfcaacagcaatccctg 2001
Db 589 TGATTAACACCTGTGTGTCTACACTTGTCTCATATAA-----ATTGACACTGCTCTGTG 642
Qy 2002 acatgtaggtcacaacttagcctctccacagcagcgtgagcagcagcagcagcagcagc 2061
Db 643 ACGTTAAGCCTCAAAACTTGTAGCATCTGTGGAGCTGCCATTCACATGTATTAATTTGCC 702
Qy 2062 tggcaactcctcagaggaagcagtcagcagcagcagcagcagcagcagcagcagcagcagc 2121
Db 703 TGGCACTGGAAGTGAAGGAGATGTGCCAGGAGCTGCAAGCACTCCCTCTGTGGCTTC 762
Qy 2122 agggacagtg 2131
Db 763 AGGCTCAGAG 772

RESULT 7
BE869819

LOCUS BE869819 700 bp mRNA linear EST 20-OCT-2000
DEFINITION 60144665JF1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850740 5',
mRNA sequence.
ACCESSION BE869819
VERSION BE869819.1 GI:10318595
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 700)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM9570 row: e column: 13
High quality sequence stop: 697.
Location/Qualifiers
1..700
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3850740"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: O190 dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 185 a 141 c 130 g 244 t
ORIGIN

Query Match 16.5%; Score 468.6; DB 10; Length 700;
Best Local Similarity 84.2%; Pred. 1e-75;
Matches 592; Conservative 0; Mismatches 94; Indels 17; Gaps 5;

QY 1106 taatattctctcggtgaacactctgacatcttaaggcgaagtgaatttcaactc 1165
|||||
Db 1 TATGATCTGCTGCTGAACCTGCTGACAAATTAAAGGGCAAAATTTTCAACTTT 60
QY 1166 gtgagttcgaagagtttttagccttaaggatgagcctcgatctctgtgtgc 1225
|||||
Db 61 ATGCTTCGCGAGAAATTTTATGACCTATAGGAATCATTTGACCTGATCTGAGC 120
QY 1226 cgcgaagattatagagctgtgattatctctgtgctgtgtcatctcacaacacac 1285
|||||
Db 121 TGTGAGCTCATGCGCTGATATTCTTTGGCGCTTCTTCTTACATATCAACAC 180
QY 1286 ccgtggagagcttttgcctctctgtgattagcatcaacacatctgtcctctgt 1345
|||||
Db 181 ACTGGGAGGCTTTGCTCTCTGATTTACATCAACCACTGGATATTCCTCTCTGTA 240
QY 1346 tgcgtgattctctctggaacaaatattctgcaacattctctctggaattgtggcc 1405
|||||
Db 241 TGTGAGTATCTCTGAGGCAATCAAAATATTGCGACATTTCCAGGAATGTTGGCC 300
QY 1406 calcatctgcagaagcttaccctgagaacacatattggaagaaatgcaaatgtttctg 1465
|||||
Db 301 CGTCATTTGATAAAGTGTGACCCCTGATACAGTGTGAGAAATGGCAAAACCGTCTCTA 360
QY 1466 catcgct-gctgtctataatgattgtgtgcatcttctctcaacatattgcgaaggtg 1524
|||||
Db 361 TATTGCTCGCTGATTAATGTTTGTGTCATTTTCTTACACATATTGCGCAAAAGTG 420

QY 1525 aagtcgcaaacctggtccatcagtcacccaagacagacaagaactgaagaaaccaataaa 1584
|||||
Db 421 AAGTACAAACTGGGCTCCTCAATGATCACCATGACACAGACACTGAGGAACCAATPAA 480
QY 1585 taatctgtctctaatatgatagtatctgtttatcatgtaacccaagtgccttgatat 1644
|||||
Db 481 TAACTCGCTCTTAAATGATATTTTATTTATCATGTAACCTCAAAAGTCCCTT-----C 535
QY 1645 tttaattgaacacatctctatcaagaataaaattgtaaaaaattgttgatt 1704
|||||
Db 536 TGTATTGTGTAAGCATTTCTATGT-----CTTTTAAATTGTAATTGATTAGATT 586
QY 1705 tgaagctgtgtaatacgaatgtaact-agtgcacataaagcaaatagctattt 1763
|||||
Db 587 T-TAAGCCCTATATCATGAAATATACCTGATGCCAGAAATATPAAATGAACGTGTT 645
QY 1764 taattatttaaccggttgcctggaacttaacattcagggtc 1806
|||||
Db 646 TACTTATGAATATATGTAAGTACGACTTCTAAGCTTACCTTACCTTACCTTACCTT 688

RESULT 8
B1860521
LOCUS B1860521 736 bp mRNA linear EST 10-OCT-2001
DEFINITION 603386787F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5',
mRNA sequence.
ACCESSION B1860521
VERSION B1860521.1 GI:16001268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM12008 row: m column: 21
High quality sequence stop: 713.
Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:5395772"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 216 a 142 c 147 g 231 t
ORIGIN

Query Match 16.2%; Score 460.2; DB 10; Length 736;
Best Local Similarity 87.4%; Pred. No. 3.4e-74;
Matches 515; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 318 acgtctcagagtggtgcagacatctgtctccataaaagttctcaacacaaaggtt 377
|||||
Db 1 ACTTCCAAAGCGCTCCAGAGCATTTCTGCTCCATTAAGATTCATCATATCAAAAGCGGT 60

QY	378	aaaagaatcccggtgggaatgcacagaactcgaagaatgatgatctctggatctctttttctcatagc	437
Db	61	AMGAAGTACCAATGGGATGCAACAACCTCAAGATGGATTCCTGGTTCCTTTTATTATGGC	120
QY	438	taacalcatacaacaattctccctctggaagatattgtgccagaagaagtgagggaagctcttg	497
Db	121	TACATCATTCACACAAATTCCTGGAGGATATGTGCCAGCAAAATAGGGGGGAAAAATGCTG	180
QY	498	ctagattctggagatctttgtctacagctatcttcacccgtttacccctccctgcctgaagt	557
Db	181	CTAGGATTTGGGATCCTTGGCACTGCTGTCTCACCCCTGTTACTCCATTCCTCTGAGAT	240
QY	558	ttcggagtcggaagccctctgtttgcaactcagagggaactagaagaaggctagaagggttgc	617
Db	241	TTAGAGATTGGACCACTCATTTACTACAGAGCATTAGAAGGACTAGAGAGGGGTGTACA	300
QY	618	tatcagacatgcatagcacatgtgctcttcacatggtctccccccttgaagaagaagcct	677
Db	301	TTTCCAGCCATGCAATGCCCCATGTGCTCTTGGGCTCCCCCTCTTGAAAGAACCAACTT	360
QY	678	ctgagtatctcatatgcacgagagacaacttggagacagtagttctctcttcctcttcgtga	737
Db	361	CTTACCATTTTCATTATGCAGAGACACAGCTTGGGACAGTAAATTTCCTTCCTTCTTGGA	420
QY	738	gtaattgctactatatagaattggactatgtctctatcttctcttggcattgttgaatc	797
Db	421	ATAAATTTCTACTAATATAGAAATGGACCTATGCTCTTCACTT-TTTGGATCATTTTGGAAAA	479
QY	798	atctggtctattctatctgacatcgcctctagttatgatacaccagaactcacaagaacatc	857
Db	480	TTTTGGTTTCCTTTGTGGATCGTGTTAGTTATGTGACACACCAACMAAACACAGAGAAATT	539
QY	858	actccctatgaaagagagatatattctcttcacatcaataaanaatcagctct	906
Db	540	TCCCAATTATGAAAGGAATACATTTCTTTCATCATATTAGAAATACAGTAT	588

RESULT	T	9
LOCUS	AM655673	
DEFINITION	106847, MARC 1BOV Bos taurus cdNA 5', mRNA sequence.	518 bp mRNA linear EST 25-APR-2001
ACCESSION	AM655673	
VERSION	AM655673.1	GI:7421499
KEYWORDS	EST.	
SOURCE	cow.	
ORGANISM	Bos taurus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.	
AUTHORS	Smith,T.P.L., Grosse,W.M., Fekling,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrentz,S.C., Bennett,G.L., Heaton,M.P., Iaegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Petree,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.	
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle	
JOURNAL	Genome Res.	11 (4), 626-630 (2001)
MEDLINE	21180013	
COMMENT	Contact: Smith TPL	

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACACTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 80 row: O column: 16
Seq primer: ATTTCAGTGACACTATAG.

```

FEATURES
  Source
    Location/Qualifiers
      1..518
      /organism="Bos taurus"
      /db_xref="taxon:9913"
      /clone_lib="MARC 1BOY"
      /tissue_type="Pooled"
      /lab_host="DH10B"
      /note="Vector: pCMV SPOR6; Site 1: XbaI; Site 2: XhoI;
      Library made from pooled tissue from lymph node, ovary,
      fat, hypothalamus, and pituitary."
BASE COUNT
  136 a 139 c 103 g 139 t 1 others
ORIGIN

```

Query Match	Similarity	15.9%	Score 451.6:	DB 9:	Length 518:
Best Local	Similarity	92.3%:	Pred. No. 1.4e-72:		
Matches	492:	Conservative	0:	Mismatches	25:
				Indels	16:
				Gaps	1:
QY	1876	gagatgagctgaataagaccctctctatacctctgtcttaattaaggtygataatactctc	1935		
DB	2	GAGATGAGCTGAATAAGACCCCTTATACCTTGGTTATTAAGGCGATATAATTCTC	61		
QY	1936	aggtcttgttaaacaatcgtttttgttaacacctctcctcaaaaattattgtcatcaagca	1995		
DB	62	AGGCTCTGTTAAACAATCGTTTTTGTACACCTCTCTCAAAAATAATTGTGTCATGACAA	121		
QY	1996	tcccgcaactgtaggtttccaacattagctctcccaaggagagtgtagcagcactgtatat	2055		
DB	122	TCCCGGACACATAGGCTCTCAAACTTTAGCTTTTACCGTAGGCGACACTGTATCAT	181		
QY	2056	tcaagctctgcaactctactctgtaggaagacatgcccagagcagctgcacatgtccctctc	2115		
DB	182	TCAGCCGCGCAACTCTCAGTGAAGGAAGCATGCCAGGAGCGTGCACACGTCCCTCTCT	241		
QY	2116	ggctcaggagagcagtgtcccaagcacttaagcagcatccaaagaccaggttcagcgcaagtc	2175		
DB	242	GGCCTTCAGGGAGACAGTGCACGACCTTAGGACGACATCCAAAGACCAAGGTCAGTCCAAAGC	301		
QY	2176	tttggtaggattctctccctctggagctgttaatgtgtgatatgaagccctgaagccaaagc	2235		
DB	302	TTTGGATGATATTCTTCCCTCTGGGCTGTAAATGATGATATAAGCCCTGAGCCCAACAGG	361		
QY	2236	gacagcggatcccaagcatcatgtttctcatcagaccctctccctctccctccagcacact	2295		
DB	362	GACACCTCGATCCCAAGTCATGATTTCATGACCCCTCTCCCTTCCCTCCAGGCGATG	421		
QY	2296	ggagtatgtcctgtgcattgaacctgcaaaagaagtgtgatgtgccttaattgacacatat	2355		
DB	422	TAAATTT-----GCAAAANGAAGATGTGATGCCTAATTATGACCATATATA	465		
QY	2356	acatcatctctgatatactactctcaatgataagagataaataatcttcaag	2408		
DB	466	ACATCATCTCTGATATCTTACTCTTCACTTGATGATAGATATAAATCTTCAAG	518		

RESULT	10
B1697765	
LOCUS	
DEFINITION	B1697765 711 bp mRNA linear EST 18-SEP-2001 603346858r1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5374407 5 ,
ACCESSION	B1697765
VERSION	B1697765
KEYWORDS	B1697765.1 GI:15660394
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 711) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
AUTHORS	Contact: Robert Strausberg, Ph.D.
JOURNAL	Email: cgapbs-re@mail.nih.gov
COMMENT	


```

QY 301 ctgccaaagataagacgctctacagagtgctcagagcattctgctccatataaagcttc 360
   ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 CTTTAAAGATATATAGACTTCACAGGCAATGTCACAGACATTCCTGCCATTAAGTTC 379
QY 361 ttcaacaacaacgggtaaaagtlaccgggtggagtgacagaacaaactcaagatgtagctcg 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 ATCATTAATCAACGGGTAGAGTACCAATGGGATGCAGAACTCAAGATGATTCCTCG 439
QY 421 gatt-ctttttctatgctacatcatcacacaaattccctggggagatattgttcacacaa 479
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 440 GTTCCCTTTTGTATGCTACATCAATCACACAGATTCCTGGAGAGATATGTTGCCACAA 499
QY 480 --agtgggggaaagcgtgttcagagatcgggattctttctaacgattccttaaccctgt 537
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 500 AATACGGGGCGGAAATGCTGCTAGATTTGGATTCCTTGCGACCTGCTGCTACCCCTGT 559
QY 538 tcaatccctcgtcgtcagatttcggagtcggagccctgttgtaactcaggagcactagaag 597
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 560 TCACTCCCATTCGTCGAGATTTAGAGCTTGACCACTCATTTGACTCAAGACACTAGAG 619
QY 598 ggtcaggaaggggtgcacatataccagccatgcattgcattggtcttcatgagcttc 654
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 620 GACTAGAGAGGGGTTCATTCAGATTCAGATTCGATTCGCTCTTGCGCTC 676

RESULT 15
BI817031 633 bp mRNA linear EST 03-OCT-2001
DEFINITION UMN07B04 Canine Brain cDNA Library Canis familiaris cDNA 5' similar
ACCESSION BI817031
VERSION BI817031.1 GI:15911742
KEYWORDS EST.
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 633)
REFERENCE Roberts, M.C., Hendrickson, J.A., Hofmann, D.E., Flickinger, G.H.,
AUTHORS Rutherford, M.S. and Mickelson, J.R.
TITLE University of Minnesota Canine Brain EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Mickelson, J.R.
Veterinary Pathobiology
University of Minnesota
1988 Fitch Avenue, University of Minnesota, St. Paul, MN 55108, USA
Tel: 612 624 1246
Fax: 612 625 0204
Email: mickel001@umn.edu
Seq primer: M13 Reverse.

FEATURES
    Location/Qualifiers
        1..633
            /organism="Canis familiaris"
            /db_xref="taxon:9615"
            /clone_lib="Canine Brain cDNA Library"
            /sex="Male"
            /note="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site_1:
            NOT1; Site_2: SalI; Tissue was taken from the frontal,
            occipital, temporal and parietal lobes, olfactory bulb,
            hippocampus, cerebellum, thalamus, hypothalamus, midbrain
            , pons, and medulla."
BASE COUNT 164 a 123 c 127 g 211 t 8 others
ORIGIN
Query Match 14.4%; Score 409.2; DB 10; Length 633;
Best Local Similarity 88.0%; Pred. No. 7.3e-65;
Matches 477; Conservative 0; Mismatches 62; Indels 3; Gaps 3;
QY 792 ggaatcattcgtgtattattatgagatcgttagtagatgatacaccagaacacaaag 851
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 GGCAATATTGGTTATTATTAGATCTTTTAGTTAGTGAACCTCCAGAGACTCACAAG 60

```

```

QY 852 acatcactcccatataaagaagatatactcttcatcatataaataacatcctctca 911
   ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ACATATCCCGCCAGAGAAAAAGATATATTCTTTCATCTATAAAAAATCATGCTTATCA 120
QY 912 cagaagtcagtcgctgtagatcactatgctgaaatcactcagcacttgggtatattgctt 971
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 CAGAACTCAGTGCATGATATACCATGCTAAATATCACTGCCGCTTGGGCTATTGTAGTA 180
QY 972 gcaaatcttcttaaacactggacttttatactattgttaaccttatgctactatcag 1031
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 181 GCACATTTTCTTACAACTGGAGACTTTTACACTTTATGACGTTATTTGGCTTACTTACATG 240
QY 1032 aaggaagtccttaaggttcacatataatcaagaagaatgggttttatactcagtcctt 1091
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AAGGACATCTTAAGGTTTCATATGTTCAAGAGATATGGCTTCTATCTGACCTTACTT 300
QY 1092 ggttgtgttatgtagatcctctgctgggtcgaagctgtgacaaatttaagggaagatg 1151
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 301 GGCTGTGGTTATGATGATCTTGTCTGCAAGCTGTGACAAATTTAAGGGCAAAATGG 360
QY 1152 aattttcaactctgtgggttcgaagaatttttaacctataaggaatgtagactgtg 1211
   ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AATTTTCAACAAATATGTCGCCAAGAGTTTATAGCTAATAGAAATGATTTGACCTGGC 420
QY 1212 atattcgtgtgcgcaagattataggtctgtgattatccttggctgttgatctccia 1271
   ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 421 GTATTCTGTGTACCCCTGTGATTTATAGTTGTGACTATTATGGCCG-TGAGGCTCTTA 479
QY 1272 accatatcaacaacccctggagggcttctcctctcgtgatttagcatcaaccatctgag 1331
   ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 ACCATATC-ACAACTGGGAGNCTTTTGCTCTTNNGGA-TTAAACATCAANCATCTGAT 537
QY 1332 at 1333
   ||
Db 538 AT 539

```

Search completed: July 15, 2002, 11:45:25
 Job time: 7560 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 ; Search time 32.66 Seconds
(without alignments)
1576.971 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836

Sequence: 1 MAAGAMTPRPYQPARPGF.....LFAKGEVQWMLNDHHGHRH 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	987	34.8	493	2	protein C38C10.2 (
2	967	34.1	472	2	hypothetical prote
3	936	33.0	560	2	brain specific Na+
4	822.5	29.0	465	2	sodium/phosphate t
5	809.5	28.5	465	2	sodium-phosphate t
6	808	28.5	563	2	probable sodium-de
7	808	28.5	576	2	protein 2K512.6 [i
8	801.5	28.3	467	2	sodium phosphate t
9	797.5	28.1	573	2	hypothetical prote
10	786.5	27.7	465	2	Na+-dependent phos
11	767.5	27.1	544	2	hypothetical prote
12	747	26.3	512	2	hypothetical prote
13	629.5	22.2	501	2	hypothetical prote
14	562	19.8	413	2	protein F2566.7 [i
15	531	18.7	537	2	hypothetical prote
16	530.5	18.7	466	2	hypothetical prote
17	492.5	17.4	561	2	hypothetical prote
18	486.5	17.2	568	2	hypothetical prote
19	476.5	16.8	592	2	hypothetical prote
20	460.5	16.2	516	2	hypothetical prote
21	452	15.9	530	2	hypothetical prote
22	445	15.7	380	2	hypothetical prote
23	444.5	15.3	543	2	hypothetical prote
24	432.5	15.3	499	2	hypothetical prote
25	431	15.2	506	2	hypothetical prote
26	419.5	14.8	493	2	hypothetical prote
27	392.5	13.8	478	2	hypothetical prote
28	389.5	13.7	445	2	hypothetical prote
29	388.5	13.7	462	2	hypothetical prote

30	386.5	13.6	473	2	hypothetical prote
31	380.5	13.4	425	2	hypothetical prote
32	372	13.1	420	2	membrane transport
33	370.5	13.1	428	2	hypothetical prote
34	362	12.8	485	2	hypothetical prote
35	360	12.7	455	1	probable glucarat
36	357.5	12.6	455	2	hypothetical prote
37	357	12.6	452	2	hypothetical prote
38	355.5	12.5	659	2	hypothetical prote
39	353	12.4	516	2	hypothetical prote
40	349	12.3	450	2	probable glucarate
41	349	12.3	450	2	probable transport
42	349	12.3	450	2	probable transport
43	339.5	12.0	445	1	hypothetical prote
44	336.5	11.9	493	2	hypothetical prote
45	334	11.8	487	2	hypothetical prote

ALIGNMENTS

Query Match	Score	DB 2;	Length	493;
Best local similarity	39.7%	Pred. No. 8.2e-71;		
Matches 194;	Conservative	94;	Mismatches 179;	Indels 22;
			Gaps 6;	
QY	55	ESTDTRPLPGAPRAEAAPVCCSARYNLALAFPGFIYALRYNLVALVWDVDSNTT	114	
DB	2	EGATTKRPLVP-----STRFALSLVMFEGCLVTYMMRTNMSFAVVCVNNKKT	49	
QY	115	LEDNFTSKACPEHSAPIYVHNHNGTKYQMDAETOGMLGSEFFVYITITDPCGVASKI	174	
DB	50	DTGVEKVSRCGEKEMPVESNSVIG-EFDWDKQKTGMVLSEFFYIGISQIIGHLSRY	108	
QY	175	GKKMLGFIIGTAVLTFTPLAADLGAGPLIVRALGIGVYFPPMHMMWSWAPPL	234	
DB	109	GKKRVFVITIGSALLTLINLPVAARTSEYALALIAIGFLOGATFPPMHMTMSWGAPL	168	
QY	235	ERSKLLISYAGAOIGTVISPLSGIICY--NMNTVFYFFFGTIGFMEFLMTWLVS	291	
DB	169	ELSVLTGTVYGAQIGNIVILPLSGFCEYFGDGPWSIFITIGFVGLWNAVMWYVSSD	228	
QY	292	TPQKKRISHYEKETIYLSLRNQSQ---KSPVWPIKLSPLMAIVAHFSTNNMTEY	347	
DB	229	KPATPRITPEKQYIVAVASMGKDKGVPSPWIKILTSPAWMACWAGHFGDWCAY	288	
QY	348	TLTLTLPTMYKEILRFNVQENGFLSLPLYSWICMLISGGAADNLRAKMNFTLCVARI	407	
DB	289	TMLVSLPFLKDVGLNLSSIGAVASTIPYIAFYLAINGVGLADLRKSGILSTLNTERRA	348	
QY	408	FSLGMITGPAVFLVAAAGFIGDYS-LAAVFLITSTLLOGFCSSGFSIHLDIAFVAGIL	466	

DB 349 AMVALIGCIFLVASGCGGDDVLIITTCGMAISGLQVAGFVNVLEIAPPSGTV 408
 QY 467 LGITNTPATIPGWGVPIAKSLTPDNTVGEMQVIFYIAAINVEGALFTFLPAKGEVQNM 526
 DB 409 MGNNTISALAGIISPAVSSYLTPNGQEMQVWLMTAIIITGALLFISFASGEVQPM 468
 QY 527 A-LNDHHGH 534
 DB 469 AKLTAEEGH 477

RESULT 2

S28286
 Hypothetical protein C38C10.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 16-Feb-1997
 C:Accession: S28286
 R:Thomas, K.
 Submitted to the EMBL Data Library, December 1992
 A:Reference number: S28285
 A:Accession: S28286
 A:Molecule type: DNA
 A:Residues: 1-472 <THO>
 A:Genetics:
 A:Cross-references: EMBL:219153
 A:introns: 50/3; 287/3; 351/3; 412/3
 C:Keywords: transmembrane protein

Query Match 34.1%; Score 967; DB 2; Length 472;
 Best Local Similarity 41.0%; Pred. No. 3.1e-69;

Matches 187; Conservative 87; Mismatches 172; Indels 10; Gaps 5;

QY 88 FFEFFIYALRVMLSVLVDVMSNTLTEDNRTSKACPEHSAPIKYHHNQTGRKKYQDAE 147
 DB 2 FFCCLITVMKRTMSPAVVCAMVNEKTDIGVEKYSRCKMTPESSSVIG-EPMDKQ 60
 QY 148 TQGMILGSFFGYIITQIPGVYASKIGKMLGFGILGTAVLTLFTPIADLVGPLYV 207
 DB 61 TTGMVLSPFYGYISQIIGGHLASRGGRVYFVITLGSALLTLNPAVARTSEVALAI 120
 QY 208 LRLEBEGEVTPPAHAMAMSSWAPLEBSKLISYAGQGLTVISLPSITCIY--- 264
 DB 121 LRAIFELQATPPAMTHMSSVMPLELSVLGVYAGQIGNVTLPLSGICEGEPD 180
 QY 265 MNTYVEYFEGTIGIFMELMWLVSDTPQKHKRISHYEKEYTLSSLRNQLSSQ---KS 320
 DB 181 GGMPSIFITIGVGLMTAVMTVSSDKPATRITPEEKQYIVTAVESMGKDTGKVP 240
 QY 321 VPMVPIILKSLPLMAIVVAHFSYMTFTLTLLPTYMKELIRNVOENGFLSLPYLSW 380
 DB 241 TPWIKLITSPAVWACMGHAGDMGAVTMLVSLPSFLKVDLGLNLSLGAVASIPYIAF 300
 QY 381 LCMILSGAANDIRAKMNSTLCVRIFSLIGMIGPAVFLVAAAGTIGCDYS-LAVAFLT 439
 DB 301 LAINAGVLAIDLRSKGLISTLNTLRAMVALIGIGIFLVASGCGGDDVLIITTC 360
 QY 440 STLGGFCGSGFSINHLIDIPSYAGILGTNTFATIPGWGVVIAKSLPDNTVGEWOT 499
 DB 361 GMAISGLQVAGFVNVLEIAPPSGTVMGNTISALAGIISPAVSSYLTPNGQEMQ 420
 QY 500 VFYIAAINVEGALFTFLPAKGEVQNM-A-LNDHHGH 534
 DB 421 VMLTAGILITGALLFISFASGEVQPMAKLTAEEGH 456

RESULT 3

I59302
 brain specific Na+-dependent inorganic phosphate cotransporter - rat
 C:Species: *Rattus norvegicus* (Norway rat)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
 C:Accession: I59302
 R:NI, B.; Rostek, P.R.; Nadi, N.S.; Paul, S.M.

Proc. Natl. Acad. Sci. U.S.A. 91, 5607-5611, 1994
 A:Title: Cloning and expression of a cDNA encoding a brain-specific Na(+)-dependent
 A:Reference number: I59302; MUID:94261635
 A:Accession: I59302
 A:status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-560 <RES>
 A:Cross-references: EMBL:U07609; NID:g507414; PIDN:AAA1646.1; PID:g507415

Query Match 33.0%; Score 936; DB 2; Length 560;
 Best Local Similarity 39.0%; Pred. No. 1.1e-66;

Matches 190; Conservative 91; Mismatches 168; Indels 38; Gaps 7;

QY 53 DGEESTDRT---PLLP---GAPRAEAPVCCSARNLALAFGFFIYALRVNLSVAL 105
 DB 38 DGRPVTHTRDPVYVCTCGRLP-----RTIILMSGLGICISFGICNIGVAI 87
 QY 106 VDMVDSNTLTEDNRTSKACPEHSAPIKYHHNQTGRKKYQDAETQGMILGSFFGYIITQI 165
 DB 88 VSMVNNST-----HRGHHVYQKQ---FNDPETHVGLIHSGFFGIVTQI 132
 QY 166 PGYVASKIGKMLGFGILGTAVLTLFTPIADLVGPLYVLRALBEGEVTPPAMA 225
 DB 133 PGGFIQKTPAANRVFGFAIVATSTLMLLPASARVHYGCVIFRILQGLVEGYTPACHG 192
 QY 226 MMSWAPLEBSKLISYAGQGLTVISLPSIGICCYMMNTYVEYFEGTIGIFMELM 285
 DB 193 IWSKNAPLEBSRLATTATPCGSYAGAVAMPPLAGVLYVSGHSVYVYVGSFGIFLYW 252
 QY 286 IWLVDTPQKHKRISHYEKEYTLSSLRNQLSSQSY---PWPPIILKSLPLMAIVVAH 340
 DB 253 LTVSYESPALHPSISEERKYEDIAIGESAKLNPVTKFNTPVRRPFTSNPVYAIIVANF 312
 QY 341 SYMTFTYTLTLTPYMKELIRNVOENGFLSLPYLSGMILCSGMILSGAANDIRAKMNS 400
 DB 313 CRSMFTYLLLSQPAVFEVEFGFISKVLGVSALPHLVMITIVIPGGIADFLSRHMS 372
 QY 401 TLCVRIFSLIGMIGPAVFLVAAAGTIGCDYSLAVAFITSTLGGFCSSGFSINHLIDIP 460
 DB 373 TTNVRKLMCGGEMATLLIVYGR-SHKGVAISFLVLAVGSGAISGFNNHLIDIP 431
 QY 461 SYAGILGINTFPATIPGWGVPIAKSLTPDNTVGEMQVIFYIAAINVEGALFTFLPAK 520
 DB 432 RVASILMGISNGVTLGSMVCPITVGMATKHKREMQVYFLASLVHGVIFVCPAS 491
 QY 521 GEVQNM-A 527
 DB 492 GEKQPM-A 498

RESULT 4

A56410
 sodium/phosphate transport protein, renal - rabbit
 N:Alternate names: sodium/phosphate cotransporter
 C:Species: *Oryctolagus cuniculus* (domestic rabbit)
 C>Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
 C:Accession: A56410; MUID:92052140
 R:Werner, A.; Moore, M.L.; Mantel, N.; Biber, J.; Semenza, G.; Murer, H.
 Proc. Natl. Acad. Sci. U.S.A. 88, 9608-9612, 1991
 A:Title: Cloning and expression of cDNA for a Na/P-1 cotransport system of kidney cor
 A:Reference number: A56410; MUID:92052140
 A:Accession: A56410
 A:status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-465 <WER>
 A:Cross-references: GB:M6466; NID:g165689; PIDN:AAA31461.1; PID:g165690
 C:Keywords: kidney; transmembrane protein

Query Match 29.0%; Score 822.5; DB 2; Length 465;
 Best Local Similarity 37.4%; Pred. No. 9.7e-58;
 Matches 172; Conservative 83; Mismatches 196; Indels 9; Gaps 4;

A: Introns: 38/3; 87/3; 224/1; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match 28.1%; Score 797.5; DB 2; Length 573;

Best Local Similarity 37.3%; Pred. No. 1.2e-55;

Matches 172; Conservative 87; Mismatches 161; Indels 41; Gaps 9;

80 RYMLALAFEGFEIVYALRVNLVSALVDVDSNTTLEDNRTSKACPEHSAPIKVHHNQTG 139

47 RMQIALAHGFAISFGISNFGVAKRMNNFT-----DAYGEVHE----- 88

140 KRYQMAETOGWILGSEFFGYITITQIPGGVASKIGKMLGFGILGTAVLTLPFPDAD 199

89 KEFFMTGTEVGMSESEFFGYAASQIPAGVIAKFAPNKLFMLGILFASLINTVTAIC-- 146

200 LGVGP-----LIVRLALEGEGEVTPPAHAMSSMAPLERSKLISYAGQLTGIS 254

147 LNFHPTDFIVAVIQWOGIALGVCYPAHMGVAKWYAPLERSKLATTTFTFGASVGMVG 206

255 LPLSGIICYMMNTVYFEFFGTIGFIMFLIMVLSDTPOKHKRISHYEKEYT---LSSD 311

207 LPAASALVSHFSMTPEFYFGALGYMSITMIVSGTSEPHGYISADEKKTITEKVGSY 266

312 RNQLSSQKSVPMVPLIKSLPLMAIYVAHFSYNTFTYLLTLPYKKEILRENVQENGL 371

267 AVKNTMLTLPWRDMTSTAVAMAIICSFORSMSFLLGNQLTVMKDYLDHIDIKNSGL 326

372 SSPLPYGSMCLMLSGQADNLRKWNSTLCYRIFSLIGMIGPAVFLVAGTIC--- 428

327 ALFPQGMCIIVTLTSGOLDYLRSSGKMSTEAARKSVNTEG-----FTVEAMVLCIAR 380

429 --DYSLAVAFLLTSTLTGFCSSGFSINHLDIAPSYAGILGTITNTPATIPGAVGVIAK 486

381 VRNPVIAVFLIACGAGAVLSGFNVNHFIDIAFRAPRIAMGLANGLAGAG-VGSIYVN 439

487 SLTPDNTVEBQTVFYIAAIVNFGAIFFTLEAKGEVQYMA 527

440 SLTYQNPBG-WQWVFLAMSIDIFIGLIFLEAKGDVLPMA 479

RESULT 10

139473

Na+-dependent phosphate cotransporter - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999

C:Accession: I39473

R: Miyamoto, K.; Tatsumi, S.; Sonoda, T.; Yamamoto, H.; Minami, H.; Taketani, Y.; Takeda, Biochem. J. 305, 81-85, 1995

A:Title: Cloning and functional expression of a Na(+)-dependent phosphate co-transporter

A:Reference number: I39473; MUID:95126933

A:Accession: I39473

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-465 <RES>

A:Cross-references: GB: D28532; NID: 9639841; PIDN: BAA05888.1; PID: 9639842

Query Match 27.7%; Score 786.5; DB 2; Length 465;

Best Local Similarity 37.7%; Pred. No. 7.1e-55;

Matches 177; Conservative 79; Mismatches 186; Indels 27; Gaps 6;

74 PCCSARYNALIAAFGEFIVYALRVNLVSALVDVDSNTT--LEENRPSKACPEHSAPI 131

11 PGCSRGVLSFLVHCNNTITRAQMLCLNTWVYVNSNDPHGLPRTSKLLDNKNKMW 70

132 KYHHNQTGKKYQMAETOGWILGSEFFGYITITQIPGGVASKIGKMLGFGILGTAVLT 191

71 -----YMSPDVQGIILSTSYGIIIVPGVYSGIYSTKMKMGFALCLSSVLS 120

192 LFPPIADIGVGLIVRLALEGEGEVTPPAHAMSSMAPLERSKLISYAGQLTG 251

121 LLIIPPAIGVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 180

QY 252 VISLPLSGIICYMMNTVYFEFFGTIGFIMFLIMVLSDTPOKHKRISHYEKEYTSSL 311

181 FIVLVTGVICESLGMPPMYIFIGACCANCLMFVLPYDDPDHDCISIGKEEYITSSL 240

QY 312 RNQLSSQKSVPMVPLIKSLPLMAIYVAHFSYNTFTYLLTLPYKKEILRENVQENGL 370

241 VQGVSSRSQSLPIKALIKSLPVAISIGSEFFWSHIMLTLYPMEINSMALHNIKENG 300

QY 371 LSLPLYSGLMCLMLSGQADNLRKWNSTLCYRIFSLIGMIGPAVFLVAGTICDY 430

301 LSLPLFAMTICGNLAGQSDFFLTRNLSIVARLFTAGFLPLPAIFGVCLPYSTSP 360

QY 431 SLAVAFLLTSTLTGFCSSGFSINHLDIAPSYAGILGTITNTPATIPGAVGVIAKSLT 489

361 YSIVIFLLIAGATGSCFLGVFINGLDIAPRYGFI-----KACSTLGMIGGLIASTLTG 416

QY 490 -----PDNTVEBQTVFYIAAIVNFGAIFFTLEAKGEVQYMA 532

417 LIKQDPESA---WFKTFILMAIINVTLGLIFLVATAEIQDAWAKEKH 462

RESULT 11

T24633

hypothetical protein T07A5.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24633

R: Buck, D.

submitted to the EMBL Data Library, February 1995

A:Reference number: T24633

A:Accession: T24633

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-544 <WILL>

A:Cross-References: EMBL:248055; PIDN:CA88134.1; GSPDB:GN00021; CESP:T07A5.3

A:Experimental source: clone T07A5

C:Genetics:

A:Gene: CESP:T07A5.3

A:Map position: 3

A:Introns: 38/3; 87/3; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match 27.1%; Score 767.5; DB 2; Length 544;

Best Local Similarity 35.9%; Pred. No. 2.8e-53;

Matches 166; Conservative 89; Mismatches 164; Indels 43; Gaps 9;

80 RYMLALAFEGFEIVYALRVNLVSALVDVDSNTTLEDNRTSKACPEHSAPIKVHHNQTG 139

47 RMQIALAHGFAISFGISNFGVAKRMNNFT-----DAYGEVHE----- 88

140 KRYQMAETOGWILGSEFFGYITITQIPGGVASKIGKMLGFGILGTAVLTLPFPDAD 197

89 REFMTGAEVGMSESEFFGYAASQIPAGVIAKFAPNKIFMGLIIVAFMNLISAIEN 148

QY 198 ----ADLGVPLVLTALRGALGEGVTPPAHAMSSMAPLERSKLISYAGQLTGYI 253

149 FHHYTDIFV---MVQAVGALAGVLYPAMHGVAKWYAPLERSKLATTAFTGSSGVMT 205

QY 254 SLPLSGIICYMMNTVYFEFFGTIGFIMFLIMVLSDTPOKHKRISHYEKEYTSSLRN 313

206 GPASAYLVSHFSMTPEFYFGALGYMSITMIVSGTSEPHGYISDEKKTITEKVG 265

QY 314 ---QLSSQKSVPMVPLIKSLPLMAIYVAHFSYNTFTYLLTLPYKKEILRENVQENGL 370

266 VAVKNNSTLTLPWRDMTSSAVAMAIICFCRSWGFLLGNQLTVMKDYLDHIDIKNSGF 325

QY 371 LSLPLYSGLMCLMLSGQADNLRKWNSTLCYRIFSLIGMIGPAVFLVAGTIC-- 428

326 ISIFPQGMCIIVTLATGOLCDYLRSSGKMSSTEAVKSVNTEG-----FTVEAMVLCIAR 379

QY 429 ---DYSLAVAFLLTSTLTGFCSSGFSINHLDIAPSYAGILGTITNTPATIPGAVGVIAK 485

380 FVRDPAVIAVTCIVACTGSGSVLSGFNVNHFIDIAFRAPRIAMGLANGLAGAVG-VGGMVT 438

QY 486 KSLFDPNTVGEWOTVFYIAAIVFGAIFFTLFAKGEYONMA 527
 Db 439 NTVTYQNPBG-WKRVFLLAMAIDIFGVIFLLIFAKGVLPRMA 479

RESULT 12

H84698
 hypothetical protein At2g29650 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: H84698
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Beato, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 Euse, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402:761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: H84698
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512 <STO>
 A:Cross-references: GB:AE002093; NID:g3582333; PIDN:AAC35230.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g29650
 A:Map position: 2

Query Match 26.3%; Score 747; DB 2; Length 512;
 Best Local Similarity 35.1%; Pred. No. 1.1e-51;
 Matches 157; Conservative 80; Mismatches 166; Indels 44; Gaps 7;

QY 80 RYNLAIAFFGFIVYALRVNLVSALVDMVDSNTTLEDNRTSKACPEHSAPIKVHHNQTG 139
 Db 102 RNVIVLFCFSFLCLNMDRVMSIALIPM-----S 131
 QY 140 KRYQMDAETQGMILGSEFYGIITQIPGCVASKIGKMLGFIGLTAVILFTPIAAD 199
 Db 132 AEYGMNPAATVGLIQSFEFGYLLTQIAGIWAQTVGKRVLFGVIMWSIATILTPVAAK 191
 QY 200 LGVGPVLYRLBELGSGVTPRAHAMSSVAPLERSKLLISYAGAOQTVISPLSG 259
 Db 192 LGLPYLLVRAKMGVGEVAVAMPANMLISKVWPQERSRLATLVSGMYLGSVGLAFSP 251
 QY 260 IICYMMNTYVFYFGTIGIFWFLIMLWVSDTPQKHRIHYKEYLLSSRLRQSSOK 319
 Db 252 FLIHQFGPVSFYFSGISGTWMLMLKAESPLEDPTLPERKLIADMCASK-EPVK 310
 QY 320 SVNPVPIKSLPLMAIVVAHFSYNMTFTLLTLLPTYKKEILRFNVQENGFLSLPYIGS 379
 Db 311 SIPRLLISKPPVALLSCHCHWNGTIFILLTWPPTYHQVLFKTNMESGILLSVFP--- 366
 QY 380 WLCMILSGQA----ADNLRAKWNSTLCVRIFSLIGIGIPAVPLVAAGFIGCQYSLAVA 435
 Db 367 WMTMAISANAGMTADTLVSR-GFSVTVNRKIMOTIGLGPAPFLTQIKHIDSP-TMAVL 424
 QY 436 FLITSTLIGCGSSGFSINHDIAPSYAGILGTTNFATIPGAVPIAKSLTPDMVTG 495
 Db 425 CMACSSQGDATASQSGISNHDIAPRYSGVLLGSLNAGVAGVIGTATATGHIIQH---G 481
 QY 496 EMQTVFYIAAIAINFGAIFFTLFAKGE 522
 Db 482 SMDDVFETISGLVGLVGVITWMLFSTGE 508

RESULT 13
 B89135
 protein F25G6.7 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: B89135
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
 A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: B89135
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-501 <STO>
 A:Cross-references: GB:chr_V; PIDN:AMC25800.1; PID:g2384837; GSPDB:GN00023; CESP:F25G
 C:Genetics:
 A:Gene: F25G6.7
 A:Map position: 5

Query Match 22.2%; Score 629.5; DB 2; Length 501;
 Best Local Similarity 31.6%; Pred. No. 2.5e-42;
 Matches 153; Conservative 101; Mismatches 191; Indels 39; Gaps 14;

QY 62 PLTPGAPRAAPVCC----SARYNLIAFFGFIVYALRVNLVSALVDMVDSNTTLED 117
 Db 6 PISDGP-SQORPLTYRFPSPWRMTSIMLFCGCVHLMNSNGMAIVCMVNSATY-D 63
 QY 118 NRTSKACPEHSAPIKVHHNQTGKRYQMDAETQGMILGSEFYGIITQIPGCVASKIGK 177
 Db 64 NET---YPENAAPL-----LDWSSDQGYIFSAFNAGLLVMLETGG-MADKRNK 109
 QY 178 MLGFGILGTAVLFT--FT-PIADLGVPLVYRALEGEGVTPFAHAMSSVAPL 234
 Db 110 YMT---LVSVALASLAFTELPMAPISYVAIFSRFLVGADALLOPAMNSLITRWFPS 166
 QY 235 ERSKLLISYAGAOQTVISPLSGIIC---YYMNTYVFYFGTIGIFWFLIMLWVS 290
 Db 167 ERSYALGLATGGRIQIGLIIIPAGALCSOTEIFGWPISFYLSGFGVLFISYIFLGA 226
 QY 291 DTPQKHRIHYKEYLLSSLRNQ----LSQSKSVNPVPIKSLPLMAIVVAHFSYMTF 346
 Db 227 DKPSQKQISDNELKFTTISQSDGVKKRERKVPKNKILSGAVALSVISLVCHPEPL 286
 QY 347 YTLTLLPTYKKEILRFNVQENGFLSLPYLGSWLMILSGQADNLR-AKNMFSTLCV 404
 Db 287 MTLNMFPLSYLHDVHNHSTENGILSLPYSLMFAIGSSYLNTWLOKNTWKKIDIC- 345
 QY 405 RRIISLIGIPAVLVAAGFIGDYS-LAVAFITSTLIGCGSSGFSINHDIAPSYA 463
 Db 346 KVLNLSIGSIGLIGVFLAAAFELDNEHAMAVLFLCLSMASAGLTPCCQALAVSVAPYS 404
 QY 464 GILGINTFATIPGMGPVIAKSLFPDMVNGEQTVFYIAAIAINFGAIFFTLFAKGEV 523
 Db 405 GAVTGFTEFFVAVSGIIPITTKMTVDRKTAEMNLVFIYSTVIAIPPIIIFNVGSTEV 464
 QY 524 QNMA 527
 Db 465 QNMA 468

RESULT 14
 T01534
 hypothetical protein A.IG005110.nm - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 17-Mar-2000
 C:Accession: T01534
 R:Andrews, S.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of A. thaliana IG005110.
 A:Reference number: Z14347
 A:Accession: T01534
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-413 <AND>
 A:Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252847
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4

A: Introns: 79/3; 115/3; 128/3; 200/3; 288/3; 331/3; 353/3; 382/1
A: Note: A_IG005110.nm
C: Superfamily: hexuronate transporter

Query Match	Score	DB	Length
19.88	562	2	413

```

OY      80 RYNLAIAFEFFETVIALRVMLSAVALDWDMSNTTLEDNRTSKACPEHSAPIKHNQGTG 139
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      61 RMVIVLLCFSEFLLCNMDRVMNSTAILRP-----S 90

OY      140 KRYOMDAETOGWILGFEEFYGYITPOLRGYAVAKIGCKMILFGILGTAVLLETPRIAD 199
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      91 OEYMSSATVGLLIOSSEFWGILLTLIIGIADKFGKK-----128#

OY      200 LGVERPLIVRLDELGESYTPRRAHAMSSNARPLESKLTISTYAGAOIGTVSLPISG 255
        |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      129 -----GVAMPRAANNMISKIPVISEKRSIALYISMYIGSYVLGAFSP 171

OY      260 IICYYMMNTVYEFYEFGTIGLFWELMTIMLVSDPFOKKRHISHYEKEYEILSLRNOLSSOK 319
        ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      172 MLTIKFGMPVSVEYSFGSLGISIMFLMLLKFAVSPKDDPDLSSEKKVYLIGSKRR -EPVT 230

OY      320 SVPMWPIKLSLPLMAIYAHHNSYWTYTLLTLPY-----MKEL 361
        :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      231 VIPKLKLSKPVAALLISHCHNMCGFILLTWMPITYNOARSASVSLTFENIFCEOV 290

OY      362 RFNVQENGFSLSLYLGSWLCMIISGOADNLRAKNWSTLCVARIFSILGMIGAFAFLV 421
        ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      291 KFNLESGLCYLP-----WLTM-----AYFAN 313

OY      422 AAGEFIGCDYSILAVALFITSTLLSGFCSSGFSINHLDIAPSAYAGILGITNTFANIPOMG 481
        |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      314 IGGWIA--DLTVSRGLSTITNGSDAFSOSGLYXSNHODIGPRAGVYLLGLSTAGVLAGVG 371

OY      482 PVIAKSLLPDNTVGEQMOTVFYIAAIAINVGAIFFTLRAGE 522

Db      372 TAATGYILOR---GSMDDVEKVAALVILIGLVNNLRATGE 409

```

RESULT 15
T45634
hypothetical protein F1312_30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45634
R:Choline, N.; Robert, C.; Brotlier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; S
submitted to the Protein Sequence Database, November 1999
A:Reference number: Z23010
A:Accession: T45634
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-537 <CHO>
A:Cross-references: EMBL:AL13392
A:Experimental source: cultivar Columbia; BAC clone F1312
C:Genetics:
A:Map position: 3
A:Insertions: 35/2; 153/3; 229/2; 350/3; 399/3; 445/3; 457/1
A:Note: F1312.30

Query Match	18.7%;	Score 531;	DB 2;	Length 537;
Best Local Similarity	27.6%;	Pred. NO. 1.9e-34;		
Matches 151; Conservative	86;	Mismatches 195;	Indels 116;	Gaps 15

```

QY      33 ASTAHGVGVMSS-----PYEDLARN-----DGEESTRTPLTPEAPR 69
      20 SSSSSLTGVIYSSFDWRIPLPERDRRKLCTGRVYNSLKTGTNTSVD---LCGIPR 75
QY      70 AEAAPVCCSARYN-----LAIIAFFGFYIALRVNLSVA 104
      11

```

[illegible]

Search completed: July 15, 2002, 08:22:58
Job time: 43 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 : Search time 17.92 seconds
(without alignments)
1158.129 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836
Sequence: 1 MAAGAMTPPPVQPARPGF.....LPAKGEVQWALNDHHGHRH 536

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	987	34.8	YLD2_CAEEL	Q03567 caenorhabdi
2	822.5	29.0	NPT1_RABIT	Q28722 o renal sod
3	809.5	28.5	NPT1_MOUSE	Q61983 mus musculu
4	808	28.5	Y006_CAEEL	P34644 caenorhabdi
5	806.5	28.4	NPT1_RAT	Q62795 rattus norv
6	801.5	28.3	NPT1_HUMAN	Q14916 homo sapien
7	767.5	27.1	YRT3_CAEEL	Q10046 caenorhabdi
8	686	24.2	NPT3_HUMAN	Q00624 homo sapien
9	615.5	21.7	NPT4_HUMAN	Q00476 homo sapien
10	486.5	17.2	YKH4_CAEEL	P34272 caenorhabdi
11	360	12.7	GUDP_BACSU	P42237 bacillus su
12	349	12.3	GUDP_ECOLI	Q46916 escherichia
13	339.5	12.0	DGOT_ECOLI	P31457 escherichia
14	326	11.5	YRBO_BACSU	P37489 bacillus su
15	318.5	11.2	EXUT_BACSU	Q34456 bacillus su
16	310	10.9	GARP_ECOLI	P42613 escherichia
17	270	9.5	EXUT_ECOLI	P74470 escherichia
18	267	9.4	YFAY_ECOLI	P74398 escherichia
19	264	9.3	YJIL_ECOLI	P33398 escherichia
20	258.5	9.1	GUDP_PSEPU	P42205 pseudomonas
21	242	8.5	TUB3_AGRVI	P71369 agrobacteri
22	236	8.3	TUB4_AGRVI	P71369 haemophilus
23	216	7.6	YB04_HAEIN	Q05181 pseudomonas
24	198	7.0	PHU1_PSEPU	P37948 bacillus su
25	194.5	6.9	GLPT_BACSU	P57057 homo sapien
26	190	6.7	GLPT_HUMAN	P08194 escherichia
27	189.5	6.7	GLPT_ECOLI	Q9P138 chlamydia m
28	172.5	6.1	YHPT_CHLMU	P33637 bacillus su
29	169.5	6.0	YHFA_BACSU	P77228 escherichia
30	169.5	6.0	YHPT_ECOLI	P33352 escherichia
31	168	5.9	YHIB_ECOLI	P94131 actinobact
32	168	5.9	MUCK_ACICA	Q10097 schizosacch
33	167.5	5.9	YAOI_SCHPO	

34	166	5.9	457	1	PCAK_ACICA	Q43975 actinobact
35	166	5.9	480	1	GLPT_HAEIN	P96335 haemophilus
36	164.5	5.8	455	1	UHPT_CHLPN	Q94739 chlamydia p
37	163	5.7	486	1	CMCT_NOCIA	Q04733 nocardia la
38	161	5.7	496	1	NANT_ECOLI	P41036 escherichia
39	160.5	5.7	403	1	MHPT_ECOLI	P77589 escherichia
40	160.5	5.7	515	1	YMT2_RAT	Q01827 rattus norv
41	160	5.6	451	1	YHAI_BACSU	P37514 bacillus su
42	160	5.6	543	1	SGEL_YEAST	P33335 saccharomyc
43	156.5	5.5	464	1	GALP_ECOLI	P37021 escherichia
44	154.5	5.4	388	1	NORA_STAMM	P21191 staphylococ
45	154.5	5.4	396	1	BCK_ECOLI	P28246 escherichia

ALIGNMENTS

RESULT	ID	YLD2_CAEEL	STANDARD:	PRT:	493 AA.
AC	003567:				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Hypothetical 52.7 kDa protein C38C10.2 in chromosome III.				
GN	C38C10.2.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
OC	Rhabditidae; Pelodierinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BRISTOL N2;				
RX	MEDLINE=94150718; PubMed=7906398;				
RA	Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,				
RA	Bonfield J., Burton J., Connell M., Copey T., Cooper J., Fraser A.,				
RA	Croxton M., Dear S., Du Z., Durbin R., Favellio A., Fraser A.,				
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,				
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,				
RA	Lattelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,				
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,				
RA	Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,				
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,				
RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,				
RA	Woldman P.;				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.				
RT	elegans."				
RT	Nature 368:32-38(1994).				
RL	[2]				
RM	REVISIONS.				
RP	Jones S.J.M.;				
RA	Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.				
CC	-1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.				
CC	-1- SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: Z19153; CNA79549.1; -				
CC	PIR: S28286; S28286.				
DR	WormPep: C38C10.2; CE08647.				
DR	InterPro: IPR003662; sub-transporter.				
DR	Pfam: PF00083; sugar tr. 1.				
KW	Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;				
KW	Sodium transport.				
FT	TRANSMEM 10 30				POTENTIAL.
FT	TRANSMEM 85 105				POTENTIAL.
FT	TRANSMEM 112 132				POTENTIAL.

FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 348 368 POTENTIAL.
 FT TRANSMEM 375 395 POTENTIAL.
 FT TRANSMEM 406 426 POTENTIAL.
 FT TRANSMEM 441 461 POTENTIAL.
 SQ SEQUENCE 493 AA; 52737 MW; 60792715D32553DD CRC64;

Query Match 34.8%; Score 987; DB 1; Length 493;
 Best Local Similarity 39.7%; Pred. No. 1,1e-64;

Matches 194; Conservative 94; Mismatches 179; Indels 22; Gaps 6;

QY 55 EESDRTPLPGAPRAEAPVCCSARYNLALAFEGFFIYALVNLSVLMVDVNDNTT 114
 DB 2 EGATTKRLVP-----STRALSLVMEFGCLVITYMMTNMNSFAVVCWVNEKKT 49
 QY 115 LEDRTSKAGEPESAPIKVHHNQTGKKYQMDAETQGMILSGFEYGIITQIPGYSASKI 174
 DB 50 DTGVEKVSRCGKEMTPVESNSSVIG-BEDMDKQTGTGNVLSFEFGYIGSQIIGHLASRY 108
 QY 175 GKKMLFGILGTVLFLTPPIAADGVPLIVRALEGEGVTPPAMAMSSMAPPL 234
 DB 109 GKKRVVETVITIGSALITLNLNPNVAARTSEYALAILRAIGFLOGATFPAMHTMMSVWGPP 168
 QY 235 ERSKLSISYAGALGTIVSLPLSGIICYY---NMNVTYVFFEGFTIGFPEFLMIVLVS 291
 DB 169 ELAVLTGVTAGAGIAGVITPLPSGLCEYFGDGGPISIFIIIGFVGLTAWMIVYSSD 228
 QY 292 TPQKKRISHEKEYILSLRNOLSSO----KSPVWPPIKSLPLMAIVAHFSYNMTFY 347
 DB 229 KPAHPRITPEEKQYIVTAEVASGKXDTGKVPSTPWIKILTSPPAVMCMAGHFGDMGAY 288
 QY 348 TLTLFLTYKKEILRFVNOENGFLSLPYLGSWLCMTLSCGADNDRAKNNEFLCYRRI 407
 DB 289 TMLVSLSPFLKDVILGLMLSSGAVASIPYATAYLAINAGVLDLTRSKILTLNTRRA 348
 QY 408 FSLIGMGPAVEFLVAGFICGDYS-LAVAFLLISTTLGSCSSGFSNLHDIAPSYAGIL 466
 DB 349 AMVALVIGGIFLVASGCGGQDVLIIFITCGMAISGLQYAGFVNVYLEIAPFESGTV 408
 QY 467 LGITNFEATIPGMVPIAKSLPDMNTVGMQTFYIAAINVGAIFTLFAKEGYQNV 526
 DB 409 MGTGNTISALAGIISPAVSSYLTPNGQEMQWMLTLAGILTGALLSIFASGEVQPW 468
 QY 527 A-LNDHHGH 534
 DB 469 AKLTAERGH 477
 RESULT 2
 NPRT_RABIT STANDARD; PRT; 465 AA.
 ID NPRT_RABIT
 AC Q28722;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate cotransporter 1) (Na(+)/Pi cotransporter 1) (Renal sodium-phosphate cotransport protein 1) (Renal Na(+)-dependent phosphate cotransporter 1) (NAP1-1).
 GN SLC12A1 OR NPRT1.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCB1_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney cortex;
 RX MEDLINE=92052140; PubMed=1946375;

RA Werner A., Moore M.L., Mantel N., Biber J., Semenza G., Murer H.;
 RT "Cloning and expression of cDNA for a Na/Pi cotransport system of
 RT kidney cortex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9608-9612(1991).
 CC -FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
 CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
 CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
 CC -SUBCELLULAR LOCATION: Integral membrane protein.
 CC -TISSUE SPECIFICITY: KIDNEY CORTEX AND LIVER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M76466; AAA31461.1; .
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
 FT TRANSMEM 79 98 POTENTIAL.
 FT TRANSMEM 109 129 POTENTIAL.
 FT TRANSMEM 171 191 POTENTIAL.
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 FT TRANSMEM 304 324 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 FT TRANSMEM 428 448 POTENTIAL.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 465 AA; 51798 MW; 29B33FB1EADDCAD CRC64;
 Query Match 29.0%; Score 822.5; DB 1; Length 465;
 Best Local Similarity 37.4%; Pred. No. 9.5e-53;
 Matches 172; Conservative 83; Mismatches 196; Indels 9; Gaps 4;
 QY 74 PVCCSARYNLALAFEGFFIYALVNLSVLMVDVNDNTTLEDNRTSKACPEHSAPIKY 133
 DB 11 PCCSFERYVLALFMHFCNIVIAQRMCLSLTNAMV-NNTNLHGS-----PMTSAEKL 63
 QY 134 HHNOTGKKYQMDAETQGMILSGFEYGIITQIPGYSASKIIRKLIGFALPLSLVSIF 133
 DB 64 -DNTRKPNVMSPDVGGIIFSSIFYGAFLLQIPVGYISGIYSIRKLIGFALPLSLVSIF 122
 QY 194 TPIAADGVPLIVRALEGEGVTPPAMAMSSWAPPLERSKILSISYAGALGTVI 253
 DB 123 IPQAAAGETWILICRVVGGITGTGTTAQHLEIWMKAPPLERGLTSMLSGFLGPI 182
 QY 254 SLPISGICYYMMWTFYFFFGTIGFVFLMLMVLVSDTPQKKRISHEKEYILSLRN 313
 DB 183 VLLVTGICLSELPMPVFFYFGACGCAVCLLMEVLYDDPKDHPVSLHEKEYITSLIQ 242
 QY 314 QLS-SKSPVWPPIKSLPLMAIVAHFSYNMFTYTLTLPLPYKKEILRFVNOENGFLS 372
 DB 243 QGSSTROSPLTKMTISLPLMAISFCFAIWMYSRLIVTTPULINSMLVHVDREGLS 302
 QY 373 SLPYISGMLCMILSGOADMIRAKWNESTLCVRIEFLSIMGIPAVFLVAGFICGYSL 432
 DB 303 SLPLFLAMICGVTAHRAIDLMSRNMSLAIKIKLTAIGLPLIYFSMKLLYLSGFS 362
 QY 433 AVAFLLISTTLGFCSSGSINHLDIAPSYAGILGTFNTEATIPGMVGPVIKSLTPDN 492
 DB 363 TIFELILANASSFCGLGALINLADLPARYVYFVTKGVTLIGTGWTSTVAGFLSQD 422
 QY 493 TVGEMQTFYIAAINVFGAIFTLFAKEGYQWMLNDH 532
 DB 423 PESSWFKIFLMSIINVIVIFLIFAKAEDIDMAEKOH 462


```

RESULT 3
NP1_MOUSE STANDARD: PRT: 465 AA.
ID NP1_MOUSE
AC 061983:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate
DE cotransporter 1) (Na(+)/Pi cotransporter 1) (Renal sodium-phosphate
DE transport protein 1) (Renal Na(+)-dependent phosphate cotransporter
DE 1).
GN SLC12A1 OR NP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=9535846; PubMed=7611445;
RA Chong S.S., Kozak C.A., Liu L., Kristjansson K., Dunn S.T.,
RA Bourdeau J.E., Hughes M.R., and expression analysis of a mouse renal
RT "Cloning, genetic mapping, and expression analysis of a mouse renal
RT sodium-dependent phosphate cotransporter."
RL Am. J. Physiol. 268:F1038-F1045(1995).
CC -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC Na+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: KIDNEY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77241; CA545459.1; -.
DR MCD: MGI:103209; Slc12a1.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSSEM 117 137 POTENTIAL.
FT TRANSSEM 176 196 POTENTIAL.
FT TRANSSEM 199 219 POTENTIAL.
FT TRANSSEM 260 280 POTENTIAL.
FT TRANSSEM 304 324 POTENTIAL.
FT TRANSSEM 337 356 POTENTIAL.
FT TRANSSEM 363 383 POTENTIAL.
FT TRANSSEM 399 413 POTENTIAL.
FT TRANSSEM 429 449 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 465 AA; 51589 MW; C67EE25A2C291EEF CRC64;

```

Query Match 28.5%; Score 809.5; DB 1; Length 465;
 Best Local Similarity 36.7%; Pred. No. 8.4e-52;
 Matches 167; Conservative 90; Mismatches 189; Indels 9; Gaps 2;

```

OY 74 PYCCSARNLAILAFECFFIYALRVNLVALVDMVDSNTLTEDNRFSKACPEHSAPIKV 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 PGFCFGRGLAILLHFCIAIIMAOQVCLINTLVAMV-----NNTGSPHLISSESYEM 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 134 HHNQTGKYQDAETQGIISFPYGIITQIPGYVASKTGKMLGFGILGTAVALTLF 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 63 LDNKNPYSNPSPIQIGLITSSVFEGMWVQAPVGYLSGITPMKRITIGSSLPFLSLMSLL 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

OY 194 TPRIADLCVGLIVLRALEGEGVTPFAPHAMSSNAPPLERSKLISYAGQLGTVI 253
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 123 IPPAAQVAGVALVICRYVLOGIAQGVSTGGEIWMKAPPLERGLTSMITLSGFVMPFI 182
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 254 SLPLSGIICYMNMTYVFFCGTIGIEFWLLMTVLVSTPQKHRIKSHYEYILSLRN 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 183 VLVSGFCIDLLGPMWFFYIGVICVLSWFLFEDDPKDHDPWMSSEKDYIISPMQ 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 314 QLS-OKSVPMVPLIKSLPLMAIVAHVSYNMTFYTLTLPTWKLRLRNVQENFLS 372
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 243 QASSGRQSLPIKAMKSLPLMAIILNSFAITWSLSLTYPTITSTYLVHNVENGLLS 302
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 373 SLPLSGIICYMNMTYVFFCGTIGIEFWLLMTVLVSTPQKHRIKSHYEYILSLRN 432
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 303 SLPLAVICGLIAGQMSDFLTRKISIVYRKLFTLISGFCVYIFIMCLLYSNYS 362
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 433 AVAPLTISTYLGFCSSGFSFNHDIAPSYAGILGITNTFATIPGAWGVPIAKSLPPDN 492
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 363 TVIEFTLANSTLFSYCCQLINALDIAPRYVGFKAVALIGMFGILSTYAGLILNOD 422
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 493 TVGEGQVFEYIAALINVGALFEFLFAKGEYQNNMA 527
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 423 PEYAMHKISFLMAGINVTCLVFLFAKGEIQDMA 457
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 4
Y006_CAEEL STANDARD: PRT: 576 AA.
ID Y006_CAEEL
AC P34644;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 52.2 kDa protein ZK512.6 in chromosome III.
GN ZK512.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabdilitida;
OC Rhabdilitidae; Pelodierinae; Caenorhabdilitis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favallo A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sultson J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z22177; CA80150.1; -.
DR PIR: S40767.
DR Wormpep: ZK512.6; CE01109.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;

```



```

RESULT 6
NPT1_HUMAN
ID NPT1_HUMAN STANDARD: PRT; 465 AA.
AC 014916; 013783;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate
cotransporter 1) (Na(+)/PI cotransporter 1) (Renal sodium-phosphate
transport protein 1) (Renal Na(+)-dependent phosphate cotransporter 1)
DE (Na/PI-4).
GN SLC17A1 OR NPT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94117004; PubMed=8288239;
RA Chong S.S., Kristjansson K., Zoghbi H.Y., Hughes M.R.;
RT "Molecular cloning of the cDNA encoding a human renal sodium
phosphate transport protein and its assignment to chromosome
6p21.3-p23."
RL Genomics 18:355-359(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RX MEDLINE=95126933; PubMed=7826357;
RA Miyamoto K.-I., Tatsumi S., Sonoda T., Yamamoto H., Minami H.,
RA Taketani Y., Takeda E.;
RT "Cloning and functional expression of a Na(+)-dependent phosphate co-
transporter from human kidney: cDNA cloning and functional
expression."
RL Biochem. J. 305:81-85(1995).
CC -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY CORTEX, LIVER AND BRAIN
CC BUT NOT IN OTHER TISSUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X71355; CAA50490.1; ALT_INIT.
DR EMBL: D28532; BAA05888.1; -.
DR MIM: 182308;
DR InterPro: IPR003662; sub-transporter.
DR Pfam: PF00083; sugar_tr. 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 35 36 RA -> LM (IN REF. 2).
FT CONFLICT 77 77 I -> V (IN REF. 2).
FT CONFLICT 207 207 G -> C (IN REF. 2).

```

```

FT CONFLICT 231 231 S -> G (IN REF. 2).
SQ SEQUENCE 465 AA; 50884 MW; 687A19F2A5C4ED16 CRC64;
Query Match 28.3%; Score 801.5; DB 1; Length 465;
Best Local Similarity 38.2%; Pred. No. 3.2e-51;
Matches 179; Conservative 78; Mismatches 185; Indels 27; Gaps 6;
QY 74 PCCSARYKLATLAFGFFIYVALRVNLSVALVDWDSMT--LEENRFSKACPEHSAPI 131
DB 11 PGFCSPRIGLSFLVHCCNVITRORACLNLTVWVWNSDTPHGLPTSTKKLLDNKNPW 70
QY 132 KVHNHGTGRKYWDDETQGMILGSEFFGYITQIPGQYASKIGKMLGFGIITAVLT 191
DB 71 -----YKMSPIQGIILISTSYGIIITQVPYSGIYSKMKMGFALCLSSVLS 120
QY 192 LFTPIADVGEPFLYLRALBEGVTPRAPHAMWSSNAPLERSKLLSISYAGQOLT 251
DB 121 LLIPTAGIGVAMVVVCRVAGCAQIVATAQFEIYVKMAPLERGRLTSMSTSGFLLP 180
QY 252 VLSLPLSGIICYYMNTVYFEPFGTIGIFWFLIMVLVSDTPQKHRIHYEKYILSSL 311
DB 181 FTVLVTGVICSLGMPWVYTFGACGAVCLIMFVLYDDPKDPCISISEKEYITSSL 240
QY 312 RNDL-SSOKSVPVPLIKSLPLMAIVVAHFSYWNFTLTLLPTMYKILRFNVQENCF 370
DB 241 VQGVSSRSQSLPIKALIKLPLVWALSISGTFPFWNSINITVTPMINSMLHNINENCF 300
QY 371 LSLPLYSGLWCLILSGQADNLRAKWNFTLCVRRIISFLIGNIGAVFLVAAGTICDY 430
DB 301 LSLPLPLFAMICGNLAGQLSDFELTRNIIISVIAVRKLFPAAGFLPAIFGVCLPYLSSTF 360
QY 431 SLAVAFILITSLTGRCSSGFSTNHDIAPSYAGILGTINFPATIRPGWAPIANSLT 489
DB 361 YSLVIFLLIAGATGSCFLGCVFLNGIDIAPIRIFGLT-----KACSTYLGIMGILASTLIG 416
QY 490 -----PDNTVGEMQTVFYIAAINVFGAIFFTLFAKGEVQWALNDH 532
DB 417 LLIKQPEEA--WFKTFILMAINVTGLIFLYIVATAIQDMARKKH 462
RESULT 7
YRT3_CAEBL
ID YRT3_CAEBL STANDARD: PRT; 544 AA.
AC 010046;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hypothetical 60.5 kDa protein T07A5.3 in chromosome III.
DE T07A5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodetrinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Buck D.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 248055; CAA88134.1; -.
DR WormPep: T07A5.3; CE01648.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;

```


OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP POSSIBLE FUNCTION.
 RA MEDLINE=98447507; PubMed=97722162;
 RX Hubbard B.K., Koch M., Palmer D.R., Babbitt P.C., Gerlt J.A.;
 RT "Evolution of enzymatic activities in the enolase superfamily:
 RT characterization of the (D)-glucarate/galactarate catabolic pathway
 RT in *Escherichia coli*.";
 RL Biochemistry 37:14369-14375(1998).
 RN [3]
 RP GENE NAME.
 RX MEDLINE=20225875; PubMed=10762278;
 RA Monterrubio R., Baldoma L., Obradors N., Aguilar J., Badia J.;
 RT "A common regulator for the operons encoding the enzymes involved in
 RT D-galactarate, D-glucarate, and D-glycerate utilization in
 RT *Escherichia coli*.";
 RL J. Bacteriol. 182:2672-2674(2000).
 CC -1- FUNCTION: UPTAKE OF D-GLUCARATE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: ABA0439.1; -
 DR EMBL: AE000362; AAC75831.1; -
 DR Ecocore: EGI3169; gndp.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Transmembrane; Transport; Complete proteome.
 FT TRANSMEM 21 41
 FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 352 372 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 FT TRANSMEM 417 437 POTENTIAL.
 SO SEQUENCE 450 AA; 49142 MW; E0599FACB1BB488 CRC64;

Query Match 12.3%; Score 349; DB 1; Length 450;
 Best Local Similarity 24.5%; Pred. No. 2.7e-18;
 Matches 115; Conservative 90; Mismatches 197; Indels 68; Gaps 16;

OY 78 SARYNALIAAFGEFFIVLALRYNLVALYDWDNSNTLEDNRKSKCPHSAPIKYNHNO 137
 DB 16 NAYYVWVWLFVTSFNVDRAFLTAGSEM----- 46
 OY 138 TGKKYQMDAETQWILGSEFFYGIITQIPGVYASKIGKMLIGILTAVALTL--FT 194
 DB 47 -AKDIGLDVGMGYVASNSMAVYVIGQIGWMLDRFGSKRYVFWSIIFLWMTLLQGV 105
 OY 195 PLADIIG-VGPLIVRALBGLGCVTPPAMNHAMWSSWADPLERSKLLISTSYACAGOLGTVI 253

DB 106 DIFSGFGIIVALLFTLRFVLGLAEDSPFGNSRIVAAMPDAQRGTAVSTFNSAQIFATVI 165
 OY 254 SLPLSGIICYNNMTYVFEFFGTIGFELTWIMLVSDTPQKHKRISHYEKEYI-----L 308
 DB 166 FAPIGMWLTHEGWSHVEFFKGLGIVSIFMLKYIHE-PNQHPCGNKKELEIYAAGGL 224
 OY 309 SGLRNQLSSQK--SVWPVI--LKSLEPLMAIVYAHFSYNTFTYLLTLPTYPEKIELR 362
 DB 225 IMDDQNTKVKVPFSYKMQGQIKQLLGRSMIGIVYIGQYICINLTYFTIWPFPVYLQVANG 284
 OY 363 FNVQENGFLSLPYIGSWICMLLSQADNLAKNNFSLCYRRIFSL-IGMIGPAV--F 419
 DB 285 MSILKAGFVAVSPAVCGFVGVLGGIISD-----W-----LMRRRTGSLNIAKRTPIVMGM 334
 OY 420 IYAAGRIGCDYS---LNAVFLITSTLGGSCSGSINHLIAP-SYAGIILGITNPPA 474
 DB 335 LLSNVAVPCNYVNVEMWIIIFMALAFPGGIGALGMNAV-MADTAPKEISGLSGGLPFNMFG 393
 OY 475 TIPGMVGPVIAKSLTPDNTVGMQVTFYIAAAINVFGAIFFTLPFAKGEVO 524
 DB 394 NISGIYTPAIGIYV--GTTGSPNGAL-IYGVVHALINVLSTVLVGDIK 440
 RESULT 13
 ID DCOT_ECOLI STANDARD; PRT; 430 AA.
 AC P31457;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE D-galactonate transporter.
 GN DGCOT OR B3691.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=93315143; PubMed=7686882;
 RA Blattner F.R., Plunkett G. III, Daniels D.L., Blattner F.R.;
 RT "DNA sequence and analysis of 136 kilobases of the *Escherichia coli*
 RT genome: organizational symmetry around the origin of replication.";
 RL Genomics 16:551-561(1993).
 RN [2]
 RP SIMILARITY TO DAL5 FAMILY.
 RA Koonin E.V.;
 RL Unpublished observations (OCT-1993).
 CC -1- FUNCTION: INTAKE OF GALACTONATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L10328; AAA62043.1; ALT_INIT.
 DR EMBL: AE000446; AAC76714.1; ALT_INIT.
 DR Ecocore: EGI1715; dgcot.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Transmembrane; Inner membrane; Transport; Complete proteome.
 FT TRANSMEM 18 38 POTENTIAL.
 FT TRANSMEM 55 75 POTENTIAL.
 FT TRANSMEM 85 105 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.

FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 310 330 POTENTIAL.
 FT TRANSMEM 338 358 POTENTIAL.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 404 424 POTENTIAL.
 SQ SEQUENCE 430 AA; 47076 MW; 86B1902777F2E461 CRC64;

Query Match 12.0%; Score 339.5; DB 1; Length 430;
 Best Local Similarity 26.3%; Pred. No. 1,3e-17;
 Matches 124; Conservative 79; Mismatches 185; Indels 83; Gaps 18;

OY 80 RYNDLAFEGFEFTVYLRVNLVALVDVDSNTLEDNRTSKACPEHSAPRIKVNHOTG 139
 DB 14 RYLVLMIFIVVICYDRANLAVA-----SAHIOEFGIT 49
 OY 140 KKYQMDAETQGMILGSEFFGYITIPGGYVASKIGKMLGFLGTAFLTPTIAD 199
 DB 50 -----KAE-MGYFSARAWLYTLCQIPGFWLDKRVGSVYTFIAIFGMSVATLFOGFRY- 102
 OY 200 LGVGPLIVLRALGELGEGVTFPRAMAMSSWAPPLERSKLTISYAGAOIGTVISLPLSG 259
 DB 103 -GLMSLGLRATIGIFEARAFPTNNKMTVSWFPERHASANGFTSGOFVLAFLTPLLI 161
 OY 260 IICYMMWTVYFEGFTIGIFWLLIMLVSDPQKHKRISHYKEYI----- 307
 DB 162 WQEMLMWHWVFIYTGIGIITSLIM-FKYQRPRLTKGISKALDYIRDGGLVDGAP 220
 OY 308 -LSSLRNOLDSQKSVPPVPIKSLPRMAIYVAHFSYMMFTTLTLPTLYKELLRNVQ 366
 DB 221 VKKRAEPLTAK--DW-KLVEHRRKLGVYIGQFAVASTLMEFLTPPNYLTQEKITATL 276
 OY 367 ENGELSLPYGSMILSGOADNLRANKNFSTLCVRRIFSL-IGMIGPAV--FLVAA 423
 DB 277 KAGMTVTPFLAAVGVLSGVAD-----LVYKGFSGFARKPIITGLLIST 326
 OY 424 GFGCDYS-----LAVAFLITSTLGFCSSGFSINHLDIAP-SYAGILGITNTF 473
 DB 327 CIMGANTYNDPMIMICMLALAFEG-----NGFASITMSLVS-SLAPRRLIGLTGVENFA 380
 OY 474 ATIGMGVPIAKSLPNDNTYGEWQVFIYAAAINVGAIFFTLPAKEGVO 524
 DB 381 GGLGIGTIVPLVGYLA--QGEGFAPALVYI-SAVALLGALSTILLY-GDVK 427

RESULT 14
 YRBO_BACSU
 ID YRBO_BACSU STANDARD; PRT; 435 AA.
 AC P37489;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Hypothetical transport protein yrbO.
 GN yrbO.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT Systematic sequencing of the 180 kilobase region of the Bacillus
 subtilis chromosome containing the replication origin.";
 RL DNA Res. 1:1-14(1994).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE PHALATE PERMEASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D26185; BAO05188.1; -;
 DR EMBL; Z99124; CAB16094.1; -;
 DR Subtilist; BG10016; yrbO.
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 KW Hypothetical protein; Transmembrane; Transport; Complete proteome.

FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 347 367 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 407 427 POTENTIAL.
 SQ SEQUENCE 435 AA; 48248 MW; 9630914D9B606208 CRC64;

Query Match 11.5%; Score 326; DB 1; Length 435;
 Best Local Similarity 25.7%; Pred. No. 1.2e-16;
 Matches 116; Conservative 78; Mismatches 193; Indels 64; Gaps 16;

OY 90 GFPIV-YALRVNLVALVDVDSNTLEDNRTSKACPEHSAPRIKVNHOTGKYQMDAET 148
 DB 29 GIITNFYDRVAISVAPAIQDS-----FHLTATE 58
 OY 149 QGMILGSEFFGYITIPGGYVASKIGKMLGFLGTAFLTPTIADLGVPLIV 208
 DB 59 LGIVFSITVYSTLMQPVGSLDREGVAVTRVGMTVMSFVTL-LAFLQKLLLYLF 116
 OY 209 RALEGLGEGVTFPRAMAMSSWAPPLERSKLTISYAGAOIGTVISLPLSGIICYMMNT 268
 DB 117 RELIGLSASAFPAASKATLMPFSPRGLANSIFDSAAFSNVIGAPLVAFLVTFDWR 176
 OY 269 YVEYFEGFTIGIFWLLIMLVSDPQKHKRISHYKEYILSSLRNOLDSQKSVPPV--PI 326
 DB 177 VAFLTGICINVL-FTIPFQWYEQPERHKRISSELYI--QKHNATITBQ-IPYGTGPL 232
 OY 327 LKSL-----PLMAIYVAHFSYMMFTTLTLPTLYKELLRNVQENFSLSLPLGSMIC 382
 DB 233 LKRLFTNRKVGMLIGFTGYGFNLLTWLPFFKHTYQMDLMSGLFAVFWLLSTIS 292
 OY 383 MIISSQADNLRANKNFSTLCVRRIFSLIGMIGPAVFLVAGFTGCDYSIAVAFITISTY 442
 DB 293 GIANGGLVGYFIKKGVPNKVRYIVIGMSGGFEL--GSI-LTNNTVAIICISIG 348
 OY 443 LGGFCSS--GFSINHLDIAP--SYAGILGITNTFATIPMGVPIAKSLTP--DNT 493
 DB 349 LAGISATAPVGSWS-AELAPISVSMLSSMVLNAN-----NLFGIIAASLTGLYLF 401
 OY 494 VGEWQVFIYAAAINVGAIFFTLPAKEGVO 524
 DB 402 TGSFTLSFLVAGFVLLGLVGY-VFVLGDVYK 431

RESULT 15
 EXUT_BACSU
 ID EXUT_BACSU STANDARD; PRT; 422 AA.
 AC O34456;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Hexonate transporter.
 GN exut.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OM of: US-09-776-865-2 to: N_Geneseq_032802:* out_format : pfs
Date: Jul 15, 2002 3:41 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xih
-O/cg92.1/USPTO.spool/US0977665/runtat_15072002_0822208_27459/app_query.fasta.1.1159
-DB=N_Geneseq_032802 -QPMF=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAPEXT=4.000 -MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2
-TRANS=human40.ctd -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTWT=pfis
-NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USBR=US0977665.@CGN1.1-220 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-776-865-2
Query Length: 536
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 331.590000

score_list:

```
Sequence      Strd Orig      ZScore      EScore len Documentation
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT:AAZ50879 + 2836.00 4877.55 2.5e-25
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:ABD10325 + 2836.00 4677.55 2.5e-25
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAF55900 + 2804.00 4823.88 2.4e-24
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAK94876 + 2734.00 4701.70 1.5e-24
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT:AAZ50875 + 2717.00 4672.73 6.3e-25
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAI58115 + 2614.00 4500.74 2.4e-24
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT:AAZ50876 + 2338.00 4015.00 2.7e-21
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAD10326 + 2338.00 4015.00 2.7e-21
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT:AAZ50880 + 2217.00 3812.87 5.0e-20
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT:AAZ50881 + 1967.00 3379.69 6.7e-18
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAH9234 + 1716.00 2937.98 2.7e-18
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:ABL18113 + 992.00 1688.15 1.1e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAI59901 + 1614.00 2773.43 3.9e-11
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAK93901 + 1417.00 2433.13 3.5e-11
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAK92364 + 1333.50 2299.05 1.0e-11
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL03769 + 996.50 1695.00 4.6e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL18113 + 992.00 1688.15 1.1e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL07417 + 992.00 1688.15 1.1e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL03766 + 984.50 1663.45 2.6e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001A.DAT:AAK52406 + 962.00 1628.63 2.3e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA1998.DAT:AAV33503 + 943.00 1598.39 1.1e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA1998.DAT:AAV33503 + 942.50 1596.80 1.4e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA1996.DAT:AAAT42064 + 942.00 1596.65 1.4e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:AAH99626 + 940.50 1597.75 1.2e-8
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL13735 + 928.50 1577.53 1.6e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL20393 + 928.50 1577.53 1.6e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL13734 + 891.00 1500.35 3.2e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL20392 + 891.00 1489.34 1.3e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL12049 + 856.00 1455.00 1.1e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL12048 + 856.00 1445.46 3.6e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL09181 + 853.50 1445.46 3.6e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL14775 + 839.50 1425.04 5.0e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL14774 + 839.50 1415.72 1.0e-6
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL12402 - 837.00 1408.18 4.3e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL07416 - 837.00 1403.93 7.5e-7
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL08217 - 822.50 1393.13 3.0e-6
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL05527 + 811.00 1371.70 4.7e-6
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL15787 + 788.00 1353.94 4.6e-6
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL26651 + 782.50 1326.61 1.5e-6
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL08216 + 779.00 1304.70 2.5e-6
```

/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL09180 + 777.00 1297.93 6.0
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA1998.DAT:AAV57909 + 776.00 1311.05 1.1e
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL26653 + 772.50 1306.33 2.0
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL05169 + 762.00 1290.32 1.6
/SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2001B.DAT:ABL20677 + 762.00 1289.74 1.7

seq_name: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/NA2000.DAT:AAZ50879

seq_documentation_block:

ID AAZ50879 standard: cDNA: 2930 BP.
XX
AC AAZ50879;
XX
DT 31-MAY-2000 (first entry)
XX
DE Full length human GBS toxin receptor (HP59) cDNA.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW pathological vascularisation; cancer metastases; angiodenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 263..1873
FT /*tag= a
FT /*product= "Human GBS toxin receptor protein"
XX
PN WO200005375-A1.
XX
PD 03-FEB-2000.
XX
PP 22-JUL-1999; 99WO-US16676.
PR 22-JUL-1998; 98US-0093843.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Helleqvist CG, Fu C;
XX
DR WPI: 2000-205377/18.
XX
DR P-PSDB: AAY45089.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX
PS Claim 3; Page 89-93; 109pp: English.

The present cDNA sequence encodes full length human GBS (group B beta
-hemolytic streptococci) toxin receptor (HP59). This sequence was cloned
by using human embryonic lung cDNA library as template. Expression vectors
comprising this cDNA can be transformed into host cells to express GBS
toxin receptor and its fragments. Detecting the receptor in tissues is
used to diagnose pathological vascularisation, e.g. for detecting cancer
metastases. GBS toxin receptors are useful for treating conditions
associated with pathological angiogenesis or neovascularisation
(specifically cancer, reperfusion injury, scarring during wound healing,
keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
neural injury), and to raise specific antibodies used for treating early
onset disease. Inhibitors of this receptor are useful for treating
pathological or hypoxia-induced endothelial cell proliferation and
migration.

Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

alignment_scores:

Quality: 2836.00 Length: 536
Ratio: 5.291 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-776-865-2 x AA250879 ..

Align seg 1/1 to: AA250879 from: 1 to: 2930

```

1  MetAlaAlaGlyAlaMetThrProProArgProValGlnProAlaArgPr 17
   |||||
263  ATGGCGGGGGGGGATGACACCGCCCGCGGTCCAGCAGCTCGGCC 312
   |||||
17  oGlyGlyPheGlyLeuSerGlyArgArgSerLeuLeuGlyValAlaAs 34
   |||||
313  CGGGGGCTTCGGGCTGTCGGGCGGCTCCCTCTCTGCGCAGGtGGGGA 362
   |||||
34  eThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 50
   |||||
363  GTACACCTGCTCAGTAGCGCTCATGAGTCTCCGGTTCGAGACCTGGCC 412
   |||||
51  ArgAsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyAl 67
   |||||
413  CGGAACGATGGCGAGAGAGCAGCAGCAGCCGCTCTTCTACCGGGGCGC 462
   |||||
67  aProArgIaGluAlaAlaProValCysCysSerAlaArgTyrAsnLeuA 84
   |||||
463  CCCAGGGGCGGAGCGCTCCAGTGTGCTGCTGCTGTTACAACTTAG 512
   |||||
84  IaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeuArgValAsn 100
   |||||
513  CAATTTGGCCTTTTGGTTGTTCTTCATGTGTATGCAATTACGTGTGAT 562
   |||||
101  LeuSerValAlaLeuValAspMetValAspSerAsnThrLeuGluAs 117
   |||||
563  CTGAGTGTGGTGTAGTGTGATGTGATGATCAATACAACTTAGAAGA 612
   |||||
117  pAsnArgThrSerLysAlaCysProGluHisSerAlaProIleLysValH 134
   |||||
613  TAAATGAACTTCCAAAGCGGTCCAGACATTTGCTCCATTAAGTTC 662
   |||||
134  ISHisAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnGly 150
   |||||
663  ATCATTAATCAAAACGGGTAGAGAAATGAGATGGCATGCGAAGCAACGA 712
   |||||
151  TrpIleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProG 167
   |||||
713  TGGATTCGCGGTCTCTTTTATGCTACATCATCACAGATTTCTCGG 762
   |||||
167  yGlyTyrValAlaSerLysIleGlyGlyLysMetLeuLeuGlyPheGlyI 184
   |||||
763  AGGATATGTGCCAGCAAAATAGGGGGGAAATCTGCTAGGATTTGGGA 812
   |||||
184  IeLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeu 200
   |||||
813  TCCTTGGGACGTGCTCCTCACCTGTTCACTCCATTGCTGCAGATTTA 862
   |||||
201  GlyValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGluG 217
   |||||
863  GGAGTTGGACCACTCATTTGCTACTCAGAGCACTAGAGACTAGAGAGGG 912
   |||||
217  yValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProPro 234
   |||||
913  TGTTCACATTTCCAGCCATGATGATGATGTGCTTTGGGCTCCCTC 962
   |||||
234  euGluArgSerLysLeuLeuSerLysSerTyrAlaGlyAlaGlnGly 250
   |||||
963  TTGAAAGAGCAAACTTCTTAGCATTTTCGATTCGAGAGCAGCAGCTTGG 1012
   |||||
251  ThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyrMetAsnTr 267
   |||||
1013  ACAGTAATTTCTCTCTCTCTTCTGGAATATTTGCTACTATATGAATG 1062
   |||||
267  pThrTyrValPheTyrPhePheGlyThrIleGlyIlePheThrPheLeu 284
   |||||
1063  GACTTAATGCTTCTTACTTTTGGTACTAATTGGAATATTTTGGTTCTTT 1112

```

```

284  eThrIleTrpLeuValSerAspThrProGlnLysHisLysArgIleSer 300
   |||||
1113  TGTGATCTGCTGTAGTAGTACACACACCAAAACAAACAGAGATTTC 1162
   |||||
301  HisTyrGlnLysGluTyrIleLeuSerSerLeuArgAsnGlnLeuSer 317
   |||||
1163  CATTAATGAAAGAAATACATTTCTTCAATCAATTAAGAAATACAGCTTCTTC 1212
   |||||
317  rGlnLysSerValProThrValProIleLeuLysSerLeuProLeuTrp 334
   |||||
1213  ACAGAGTCACTGCCGTGGTGGTACCATTTTAAATCCCTGCACTTTGGG 1262
   |||||
334  IaIleValAlaHisPheSerTyrAsnTrpThrPheTyrThrLeuLeu 350
   |||||
1263  CTATCGTAGTGCACACTTTTCTTACAACTGAGCTTTTATACATTATG 1312
   |||||
351  ThrLeuLeuProThrTyrMetLysGluIleLeuArgPheAsnValGlnG 367
   |||||
1313  ACATTATTTGCTCTACTTATATGAAAGAGATCTTAAGCTTCAATGTTCAGA 1362
   |||||
367  uAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeuGlyMet 384
   |||||
1363  GAATGGGTTTTATCTTCATTTGCTTATTTAGGCTTTGTTATGTATGA 1412
   |||||
384  IeLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
   |||||
1413  TCGTGTCTGGTCAGACGTGTCAGCAATTTAAGGCAAAATGGAATTTTCA 1462
   |||||
401  ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
   |||||
1463  ACTTTATGTGTCCAGAAATTTTATGCTTATAGCAATGTATGGACCTGCG 1512
   |||||
417  aValPheLeuValAlaIaGlyPheIleGlyCysAspTyrSerLeuLay 434
   |||||
1513  AGATTCCTGTGTAAGCTGTGCTTCATTTGCTGTGATTAATCTTTGGCGG 1562
   |||||
434  aAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
   |||||
1563  TTGCTTTCCATACATATATCAACAACTGAGGAGGCTTTTCTGCTTCGGA 1612
   |||||
451  PheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeuLe 467
   |||||
1613  TTTAGCATCAACCATCTGATATGCTCTTCCTGATATCTGTAATCCCTCT 1662
   |||||
467  uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValI 484
   |||||
1663  GGGCATTCACAAATACATTTGGCACATTTCCAGGAATGTTGGGCCCTCA 1712
   |||||
484  IeAlaLysSerLeuThrProAspAsnThrValGlyGluTrpGlnThrVal 500
   |||||
1713  TTGCTAAAGTCTGACCCCTGATATACACTGTTGGAGAAATGGCAACCGTG 1762
   |||||
501  PheTyrIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrLe 517
   |||||
1763  TTTCTATATTTGCTGCTCTATTAAATGTTTGGTCCATTTTCTTTTAACT 1812
   |||||
517  uPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisHisGlyH 534
   |||||
1813  ATTGCGCAAAAGGTGAAGTACAAACTGGGCTCTCAATGATACACATGAGC 1862
   |||||
534  ISArgHis 536
   |||||
1863  ACAGACAC 1870

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAAD10325

seq_documentation_block:

ID AAAD10325 standard; DNA; 2930 BP.

XX

AC AAAD10325;

XX

DT 16-OCT-2001 (first entry)

XX Human group B beta-haemolytic Streptococci toxin receptor (HP59) DNA.
XX
XX
KW Human: group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;
KW cytostatic; vulnerrary; antithrombotic; osteopathic; vasotropic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
XX vaccine; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 263..1873
FT /tag= a
FT /product= "Human GBS toxin receptor protein (HP59)"
XX
XX WO200156598-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 02-FEB-2001; 2001WO-US03662.
XX
XX PR 02-FEB-2000; 2000US-0179870.
XX
XX PA (UYVA-) UNIV VANDERBILT.
XX
XX PI Helieryist CG;
XX
XX DR WPI; 2001-488844/53.
XX DR P-PSDB; AAE06518.
XX
XX PT Preventing or attenuating pathoangiogenic conditions e.g. cancer, by
XX PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
XX PT administering group B beta-hemolytic Streptococci toxin receptor or its
XX PT fragment
XX
XX PS Disclosure; Page 41-44; 52pp; English.
XX
XX CC The present sequence is a DNA encoding full length human group B beta-
XX CC haemolytic Streptococci (GBS) toxin receptor protein, HP59. The present
XX CC invention relates to a method for preventing or attenuating a patho-
XX CC angiogenic condition in a mammal which comprises administering to the
XX CC mammal one or more GBS toxin receptors or their immunogenic fragments
XX CC to induce or maintain an immune response to one of GBS toxin receptors.
XX CC The method is useful for preventing or ameliorating pathoangiogenic
XX CC conditions such as cancer, scarring during wound healing, gliosis
XX CC during repair of nerve injury, chronic wounds, keloids, reperfusion
XX CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
XX CC psoriasis in mammals. The proteins of the invention are also used
XX CC as vaccines.
XX
XX Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

alignment_scores:
Quality: 2836.00 Length: 536
Ratio: 5.291 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-776-865-2 x AAD10325 ..

Align seg 1/1 to: AAD10325 from: 1 to: 2930

1 MetAlaAlaGlyAlaMetThrProArgProValGlnProAlaArgPr 17
|||||
263 ATGCGCGCGGGGCGGATGACACCGCCCGCGGTCCAGCCAGCTCGGCC 312
17 OG1GlyPheGlyLeuSerGlyArgArgSerLeuLeuGlnValAlaAs 34
|||||
313 CGGGGGCTGCGGCTGCGGGCGGCGCTCCCTCTCTGCGCAGGTGGCGA 362

34 eRThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 50
|||||
363 GTACACCTGCTCAGCTAGGGCTCATGAGGTCTCCGGTTCCGAGACTGCC 412
51 ArgAsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyAl 67
|||||
413 CGGAACCATGGCGGAGAGACGACGACCGCACGGCTCTTCACCGGGGCC 462
67 aProArgAlaGluAlaProValGlyCysSerAlaArgGlyrAsnLeu 84
|||||
463 CCCAGGGCGGAGCCGCTCCAGTGTGCTGCTGCTGCTTACCAACTTAG 512
84 LaIleLeuAlaPhePheGlyPhePheIleValTYrAlaLeuArgValAsn 100
|||||
513 CAATTTGGCCTTTTGGTTCTTCATGTTGTAATGCAATTCAGTGTGAT 562
101 LeuSerValAlaLeuValAspMetValAspSerAsnThrThrLeuGluAs 117
|||||
563 CTGAGTGTGGCTTAGTGATATGATGATCAATCAACTTACACTTAGAGA 612
117 PASArgThrSerLysAlaCysProGluHisSerAlaProIleLysValH 134
|||||
613 TAATAGAACTTCCAAAGCGGTCCAGACATTCGTCCCTAAAGCTTC 662
134 IShISAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGlnGly 150
|||||
663 ATCATTAATCAAAACGGGTAGAGATCAATAGGATGCGAAGAACTCAAGA 712
151 TrpIleLeuGlySerPhePheTyrGlyTYrIleIleThrGlnIleProG 167
|||||
713 TGAATCTCGGCTCTCTTTTATGCTACATCATCATCAACAGATTCCTGG 762
167 YGlyTYrValAlaSerLysIleGlyGlyLysMetLeuLeuGlyPheGly 184
763 AGGATATGTTGCCAGCAAAATAGGGGGGAAATCTCTGAGATTGGGA 812
184 ILeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeu 200
|||||
813 TCCTTGGCAGCTGCTGCTCTCACCCTGTTCCACTCCCATTTGCTGCAGATT 862
201 GlyValGlyProLeuIleValLeuArgAlaLeuGlnGlyGlnGly 217
|||||
863 GGAAGTGGACACCATTCATGTACAGAGCACTAGAAAGACATGAGAGAGG 912
217 YValThrPheProAlaMetHisAlaMetTrpSerSerThrAlaProPro 234
|||||
913 TGTACATTTCCAGCCATGATGCCATGTGCTTCTTGGGCTCCCTCC 962
234 eugIuArgSerLysLeuLeuSerLysSerTYrAlaGlyAlaGlnGly 250
|||||
963 TTGAAAGAGCAAACTTCTTAGCATTTCTGATGAGAGACAGCAGCTTGGG 1012
251 ThrValIleSerLeuProLeuSerGlyIleIleCysTYrTYrMetAsnTr 267
|||||
1013 ACAGTAATTTCTCTCTCTCTTCTGGAATTAATTGCTACTATATGAAATTG 1062
267 pThrTYrValPheTYrPhePheGlyThrIleGlyIlePheThrPheLeu 284
|||||
1063 GACTTATGCTCTTACTTCTTCTTGGTACTATTTGGAAATTTTGGTTCTTT 1112
284 eutRPIleThrLeuValSerAspThrProGlnLysHisLysArgIleSer 300
|||||
1113 TGTGATCTGCTTAGTATGATGACACACACAAAACACAAAGAAATTGCC 1162
301 HisTYrGlnLysGlnTYrIleLeuSerSerLeuArgAsnGlnLeuSer 317
|||||
1163 CATATGAAAGCAATCATTTCTTCAATCAATTAAGAAATAGAGCTTCTTCC 1212
317 rGlnLysSerValProThrValProIleLeuLysSerLeuProLeuTrp 334
|||||
1213 ACAGAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1262
334 LaIleValAlaIleHisPheSerTYrAsnThrPheTYrThrLeuLeu 350

```
|||||
1263 CTATGAGTAGTGACACATTTCTTCAACCTGACATTTTAACTTATTG 1312
XX
351 ThrLeuLeuProThrTyrMetLysGluLeuArgPheAsnValGlnG 367
XX
1313 ACATTATTGCCACTTATATGAGAGATCCTAAGGTTCAATGTTCAAGA 1362
XX
367 uAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeuCysMet 384
XX
1363 GAATGGTTTTTATCTCATTCATGCCATTATTAGGCTCTGTGTTATGTAGA 1412
XX
384 LeuLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
XX
1413 TCCTGTCTGGTCAAGCTGCTGACAAATTAAAGGCAAAATGGAATTTTCA 1462
XX
401 ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
XX
1463 ACTTATATGTTGTCGAGAAATTTTGGCTTATAGAAATGATGGACCTGC 1512
XX
417 AlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaVal 434
XX
1513 AGTATTCTGTGAGCTGCTGGCTTCATTCGCTGATTTATCTTTGGCCG 1562
XX
434 AlaValPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
XX
1563 TTGCTTCTCTAATATATCAACACACATGGAGGCTTTGCTCTTCTGGA 1612
XX
451 PheSerIleAsnHisLeuAspIleAlaIleProSerTyrAlaGlyIleLeu 467
XX
1613 TTTACATCAACACATCTGATATTCCTCTGATGCTGATGCTGATCTCT 1662
XX
467 uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValI 484
XX
1663 GGGCATCAACAATACATTGCGCACATTCACAGAAATGGTGGCCCGCA 1712
XX
484 LeuAlaLysSerLeuThrProAspAsnThrValGlyIleTrpGlnThrVal 500
XX
1713 TTGCTAAAGTCTGACCCCTGATTAACACTGTTGGAGAAATGGCAACCGTG 1762
XX
501 PheTyrIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrIle 517
XX
1763 TTCCTATATGCTGCTGCTATTAATGTTTGGTGCATTTCTTTACACT 1812
XX
517 uPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisGlyH 534
XX
1813 ATTCCCAAGGTGAAGTACAAAACCTGGGCTCTCAATGATCACCCATGGAC 1862
XX
534 IsArgHis 536
XX
1863 ACAGACAC 1870
XX
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF55900
seq_documentation_block:
ID AAF55900 standard; DNA: 2512 BP.
XX
AC AAF55900;
XX
XX 18-APR-2001 (first entry)
XX
DE Human AST coding sequence.
XX
XX Human; AST; noctropic; immunotropic; gene therapy; Salla disease;
XX anion and sugar transporter; anion-cation symporter;
XX sialic acid transporter; ss.
XX
OS Homo sapiens.
XX
PN EP1069184-A1.
XX
PD 17-JAN-2001.
XX
PF 16-JUL-1999; 99EP-0202341.
```

```
XX
PR 16-JUL-1999; 99EP-0202341.
XX
PA (ALKU) AKZO NOBEL NV.
XX
DR WPI; 2001-193090/20.
XX
DR P-PSDB; AAB6967.
XX
PT New human transporter gene implicated in Salla disease and lysosomal
PT sialic acid transport, useful in assays for identifying new drugs, or
PT diagnosing sialic acid transport defects related to mutations in the
PT transporter gene.
XX
PS Claim 2; Page 12-13; 20pp; English.
XX
CC The present sequence is the coding sequence for human Anion and Sugar
CC Transporter (AST) protein. AST has significant homology with several
CC members of the anion-cation symporter (ACS) family of transporters. AST
CC is implicated in Salla disease, and is useful in screening assays for
CC identifying new drugs. Compounds identified via AST screening is useful
CC for preparing a pharmaceutical suitable as an activator or inhibitor of a
CC sialic acid transporter protein. The pharmaceutical may be used in sialic
CC acid associated diseases and CNS/immune related disorders.
XX
SQ Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other;
```

```
alignment_scores:
Quality: 2804.00 Length: 535
Ratio: 5.261 Gaps: 1
Percent Similarity: 99.626 Percent Identity: 99.252
```

```
alignment_block:
US-09-776-865-2 x AAF55900 ..
```

```
Align seg 1/1 to: AAF55900 from: 1 to: 2512
```

```
4 GlyAlaMetThr.....ProProArgProValGlnProAlaArgProG 18
|||||
151 GGGGGGGGTGTCATCGCCCCCGCCCGGTCAGCCAGCTCGGCCGG 200
PS
18 yGlyPheGlyLeuSerGlyArgArgSerLeuLeuCysGlnValAlaSer 35
|||||
201 GGGCTTCGGGCTGTCGGCGCGCGCTCCCTCTCTGCGCAGGTGGCAGATA 250
35 hProAlaHisValGlyValMetArgSerProValArgAspLeuAlaArg 51
|||||
251 CACCTGCTCACGTAGGCTCATGAGGTCCCGGTTCCAGACCTGGCCGG 300
52 AsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyAlaPr 68
|||||
301 AACGATGGCGAGGAGAGCAGCAGCAGCGCCTCTTACAGGGGCGCCCC 350
68 ArgAlaGluAlaAlaProValCysCysSerAlaArgTyrAsnLeuAla 85
|||||
351 ACGGGCCGAAACCCGCTCCAGTGTGCTGCTGCTGCTTACCAACTTAGCA 400
85 LeuLeuAlaPhePheGlyPhePheIleValTyrAlaLeuArgValAsnLeu 101
|||||
401 TTTTGGCCTTTTGGTTTCTTCATTTGTAATGCAATTAAGTGAATCTG 450
102 SerValAlaLeuValAspMetValAspSerAsnThrThrLeuGluAspAs 118
|||||
451 AGTGTGGCTAGTAGATATGTGATCAATCAATCAACTTAGAAGATAA 500
118 ArgThrSerLysAlaCysProGluHisSerAlaProIleLysValHis 135
|||||
501 TAGAACTTCCAAAGGCTGTCCAGAGCATTTGCTCCATAAAGTTTCATC 550
135 IsAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnGlyTrp 151
|||||
551 ATAATCAACGGGTAAAGATACCAATGGGATGCAAAATCAAGAGATGG 600
```

```

1501 AGCATCAACCATCTGGATATTGCTTCCGTATCGTCGAGG 1550
468 yLleThraSnThrPhealArthrIleProglYmeValGIyProvalILlea 485
1551 CATCAACAATGACTTTGGCAGTAATTCACAGAAATGGTGAGGCCGCATATG 1600
465 lalySserLeurThrProAspAsnThrValGIylutIrgrInTrValphe 501
1601 CTAAGAAGCTGACCCTGATACACTCTGTGGAGAAATGGCAACCGTGTTTC 1650
502 TyrILleAlaIalIeasnlVPhesGLYAlaIllephepethrlreuph 518
1651 TATATGCTGCTGCTAATTAATGTTTTGGTGCCATTTCTTACACATAAT 1700
518 eAlalySGlyGUValGlnsnTrPALaleuAsnAspHisHISGLYHIsa 535
1701 CGCCAAGTAGTACATACAAAACCTGGCTCTCATATGATCACATGACACACA 1750
535 rghIs 536
|||||
1751 GACAC 1755

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK94876
seq_documentation_block:
ID AAK94876 standard; cDNA: 2712 BP.
XX
AC AAK94876;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 4068.
XX
Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
X Homo sapiens.
OS
XX
PN EP1130094-A2.
PD
05-SEP-2001.
XX
07-JUL-2000; 2000EP-0114089.
XX
PF 08-JUL-1999; 99GP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
P-PSDB; AAM93914.
PT
830 Primers useful for synthesizing full length cDNA clones and their
use in genetic manipulation -
XX
Claim 8; SEQ ID NO 4068; 1380bp + sequence listing; English.
XX
The invention relates to primers for synthesising full length cDNA
clones. 830 cDNA molecules encoding a human protein have been
isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
molecules have been determined. Primers for synthesising the full length
cDNA are useful for clarifying the function of the protein encoded by
the cDNA. The full length clones were obtained by construction of full
length enriched cDNA libraries that were synthesised by the oligo-capping
method. The primers enable the production of the full length cDNA easily
without any special methods. The present sequence is a full length
human cDNA of the invention.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in CD-ROM format directly from EPO.
```

```

1501 AGCATCAACCATCTGGATATTGCTTCTTGATCTGGTATCTCTCGGG 1550
1502 yLleThraSnThrPhealArhrlleProglYmeTvaIGlyProvaIIleA 485
1551 CATCAACAATGATTTGGCTACTATTCCAGGAATGGTTGGCGCCGCATTTG 1600
1601 lalYSerleuThrProASPasnThrValgluTlrPglntThrValPhe 501
1651 CTAAAGCTGACCCCTGATACACTCTGTGGAGAAATGGCAACCGTGTTC 1650
1701 TyrIleAlaIaIaIleasnValPheGlYAlaIlePhePheThreuph 518
1751 TATATGCTGCTGCTATTAAATGTTTGGTGCCATTTCTTACACATTT 1700
518 eAlaYSglYgluValGlnasnTrPAlaLeuAsnASPHisIsgLYHisA 535
1701 CGCCAAAGTGAATACAAAACTGGCTCTCAATGATCACCATGACACACA 1750
535 rghIS 536
|||||
1751 GACAC 1755

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK94876
seq_documentation_block:
ID AAK94876 standard; cDNA: 2712 BP.
AC AAK94876;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 4068.
XX
Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
PD 05-SEP-2001.
XX
07-JUL-2000; 2000EP-0114089.
XX
08-JUL-1999; 99GP-0194486.
XX
11-JAN-2000; 2000JP-0118774.
XX
02-MAY-2000; 2000JP-0183765.
XX
(HELI-) HELIX RES INST.
XX
PA
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
P-PSDB; AAM93914.
XX
830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
Claim 8; SEQ ID NO 4068; 1380bp + sequence listing; English.
XX
XX
The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC
Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
CC

```

XX
SQ Sequence 2712 BP: 677 A; 619 C; 619 G; 797 T; 0 other;

alignment_scores:

Quality: 2734.00 Length: 556
Ratio: 5.238 Gaps: 1
Percent similarity: 93.885 Percent identity: 93.885

alignment_block:

US-09-776-865-2 x AAK94876 ..

Align seg 1/1 to: AAK94876 from: 1 to: 2712

```

15 AlaArgProGlyGlyPheGlyLeuSerGlyArgArgSerLeuLeuGly 31
|||||
2 GGTGGGGCCGGGGGCTTGGGGCTGTGGGGCCGGCCCTCCCTTCTGCCA 51
|||||
31 nValAlaSerThrProAlaHisValGlyValMetArgSerProValArg 48
|||||
52 GGTGGCGAGTACACCTGCTCAGGTAGGGGTGATGAGTCTCGGTTGAG 101
|||||
48 spleuAlaArgAsnAspGlyGluSerThrAspArgThrProLeuLeu 64
|||||
102 ACCGTGGCCCGAAGCATGGCGAGAGACAGCGACCGACGCGCTCTCTCA 151
|||||
65 ProGlyAlaProArgAlaGluAlaAlaProValGlyCysSerAlaArgTy 81
|||||
152 CCGGGCGCCCGCGGGCCGAGACCGCTCCAGTGTCTGCTGCTGCTTA 201
|||||
81 rAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeu 98
|||||
202 CAACTTACCAATTTTGGCCTTTTGGTTTCTTCAATGATGATGATAC 251
|||||
98 rGlyAlaAsnLeuSerValAlaLeuValAspMetValAspSerAsnThr 114
|||||
252 GTGTAAATCTGAGTGTGCTGATGATGATGATGATGATGATGATGAT 301
|||||
115 LeuGluAspAsnArgThrSerTyrAlaCysProGluHisSerAlaProI 131
|||||
302 TTAGAAAGTAAATAGAACTTCCAGAGCGTGTCCAGAGCATTTCTGCTCC 351
|||||
131 eLysValHisHisAsnGlnThrGlyLysTyrGlnThrAspAlaGlu 148
|||||
352 AAAAGTTCATCATATCAACGGGTAGAGATGACCAATGGATGACGAAA 401
|||||
148 hArgGlyTyrPheLeuGlySerPhePheTyrGlyTyrIleIleThrGln 164
|||||
402 CTCAGGAGATGATCTCGGTTCTCTTTTATGGCTACATCATCACAGAC 451
|||||
165 IleProGlyTyrValAlaSerSerIleGlyGlyLysMetLeuLeuG 181
|||||
452 ATTCTGGAGATATGTTGCCAGCAAAATAGGGGAAATGCTGCTAGG 501
|||||
181 yPheGlyIleLeuGlyThrAlaValLeuThrLeuPheThrProIleAla 198
|||||
502 ATTTGGATCCCTGGCACTGCTGCTCCACCCCTGTTCACTCCCATTTG 551
|||||
198 IaAspLeuGlyValGlyProLeuIleValLeuArgAlaLeuGluLeu 214
|||||
552 CAGATTGTGGAGTTGGACCATCATTTGATCTCAGAGCACTAGAAAGACT 601
|||||
215 GlyGluGlyValThrPheProAlaMetHisAlaMetTrpSerSerTrp 231
|||||
602 GGAGAGGGGTTCATTTCAGCCCATGTCATGCGATGTGCTCTCTGGGC 651
|||||
231 aProProLeuGluArgSerTyrLeuLeuSerIleSerTyrAlaGlyAla 248
|||||
652 TCCCTCTTGAAGAAGCAAACTCTTAGCATTTTCATATGAGAGAGAC 701
|||||
248 ILeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyr 264
|||||
702 AGCTTGGAGCAGTAATTTCTCTCTTCTGGAATTAATTTGCTACTAT 751

```

```

265 MetAsnTrpThrTyrValPheTyrPhePheGlyThrIleGlyIlePheTr 281
|||||
752 ATGAAATGAGACTTATGCTCTTCTACTTTTGGTACTATATGAAATTTTG 801
|||||
281 pPheLeuLeuThrPheTrpLeuValSerAspThrProGlnLysHisLys 298
|||||
802 GTTCTTTTGTGGATCTGGTTAGTAGTACACACACAAAAACACAGA 851
|||||
298 rGlyIleSerHisTyrGluLysGluTyrIleLeuSerSerLeuAsn... 313
|||||
852 GAATTTCCATTATGAAAGAAATACATCTTTCATCATTAAGAAATAG 901
|||||
313 ..... 313
902 GGTGAGTGCAGTACATGATCTTGGCTCAGCTGACGCCCTCCCTGCTC 951
|||||
314 .....G 314
952 CCGGGTTCAGGCATCTCCACCTCAGCTTACCAAGTAGTTGGAGCAGC 1001
|||||
314 ILeuSerSerGlnLysSerValProTrpValProIleLeuLysSerLeu 330
|||||
1002 AGCTTCTTCAAGAAATGACGTGCGGTGATACCATTTTAAATCCCTG 1051
|||||
331 ProLeuTrpAlaIleValAlaHisPheSerTyrAsnTrpThrPheTy 347
|||||
1052 CCACCTTGGGCTATCGTATGTTGACACACTTCTTTCACCACTGACTTT 1101
|||||
347 rThrLeuLeuThrLeuLeuProThrTyrMetLysGluIleLeuArgPhe 364
|||||
1102 TACTTATTTGACATATATGCTACTTATATGAAAGAGATCTCAAGGTCA 1151
|||||
364 snValGlnGluAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrp 380
|||||
1152 ATGTTCAAGAGAAATGGTTTATCTTCATTGCTTATTTAGGCTCTGG 1201
|||||
381 LeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLys 397
|||||
1202 TTATGTATGATTCCTGTCTGTCAGCTCTGACAAATTTAAAGGCAAA 1251
|||||
397 pAsnPheSerThrLeuCysValArgArgIlePheSerLeuIleGlyMet 414
|||||
1252 GAATTTTCAACTTATGTGTCGAGAAATTTTATGCTTATAGGAATGA 1301
|||||
414 IeGlyProAlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyr 430
|||||
1302 TTGACCTGCAGATATCTGTGAGCTGCTGCTTCAATTTGCTGATAT 1351
|||||
431 SerLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheC 447
|||||
1352 TCTTTGGCCGTTGCTTCTTCACTATATCAACACACACGAGGCTTTG 1401
|||||
447 sSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSerTyrAla 464
|||||
1402 CTCCTGTGATTTAGCATCAACCATCTGATATGCTCTCTCGATAGCTG 1451
|||||
464 IyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetVal 480
|||||
1452 GTATCTCTCTGGCATCAAAATACATTTGCCACTATTTCCAGGATGTT 1501
|||||
481 GlyProValIleAlaLysSerLeuThrProAspAsnThrValGlyGlu 497
|||||
1502 GGGCCCGTCATTTGCTAAAGTCTGACCCCTATTAACACTGTGGAGAA 1551
|||||
497 pGlnThrValPheTyrIleAlaAlaAlaIleAsnValPheGlyAlaIle 514
|||||
1552 GCAAAACGTGTATATATGCTGCTGATTAATGTTTGTGTCATTT 1601
|||||
514 hePheThrLeuPheAlaLysGlyValGlnAsnTrpAlaLeuAsnAsp 530
|||||
1602 TCTTTACACTATTTGCGCAAGAGTGAAGTAAACACTGGGCTCTCAAT 1651

```


531 HishisGlyHisArgHis 536
|||||
1652 CACCATGACACAGACAC 1669

seq_name: /stddsl/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAZ50875

seq_documentation_block:

ID AAZ50875 standard: cDNA; 2602 BP.

AAZ50875;

31-MAY-2000 (first entry)

Partial human GBS toxin receptor (HP55) cDNA.

Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
pathological vasculatation; cancer metastases; angiodenesis;
neovascularisation; reperfusion injury; scarring; keloid;
chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
endothelial cell proliferation; antibacterial; anticancer;
anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 58..1545

FT /*tag= a

FT /product= "Partial human GBS toxin receptor"

WO200005375-A1.

03-FEB-2000.

22-JUL-1999; 99WO-US16676.

22-JUL-1998; 98US-0093843.

(UYVA-) UNIV VANDERBILT.

Hellerglyst CG, Fu C;

WPI: 2000-205377/18.

P-PSDB: AAY45087.

New polynucleotide encoding mammalian receptor for streptococcus toxin,
useful for diagnosis and treatment of, e.g. pneumonia in neonates -
Claim 3; Page 77-80; 109pp; English.

The present cDNA sequence encodes partial human GBS (group B beta
-haemolytic streptococci) toxin receptor (HP55). This sequence was cloned
by using human embryo lung cDNA library as template. Expression vectors
comprising this cDNA can be transformed into host cells to express GBS
toxin receptor and its fragments. Detecting the receptor in tissues is
used to diagnose pathological vasculatation, e.g. for detecting cancer
metastases. GBS toxin receptors are useful for treating conditions
associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
neural injury), and to raise specific antibodies used for treating early
onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
migration.

Sequence 2602 BP; 672 A; 576 C; 583 G; 771 T; 0 other:

alignment_scores:

Quality: 2717.00

Ratio: 5.286

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-776-865-2 x AAZ50875 ..
Align seg 1/1 to: AAZ50875 from: 1 to: 2602

23 SerGlyArgArgSerLeuLeuCysGlnValAlaSerThrProAlaHisVal 39
|||||
1 TCGGGCCGGCGCTCCCTTCCTGCGAGGTGGCGAGTACACCTGCTCAGT 50
39 IglValMetArgSerProValArgAspLeuAlaArgAsnAspGlyVal 56
|||||
51 AGCGTCATGAGGCTCCGGGTGAGACCTGGCCGGGAGAGTGGCGAGG 100
56 IserThrAspArgThrProLeuLeuProGlyAlaProArgAlaGluAla 72
|||||
101 AGAGCAGCAGCAGCGCCCTTCCTACCGGGGCCGCCAGCGGGCGAGACC 150
73 AlaProValCysSerAlaArgTyrAsnLeuAlaLeuAlaPhePhe 89
|||||
151 GCTCCAGTGTGCTGCTCTGCTGCTACACTGACATTTTGGCTTTT 200
89 eglPhePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeu 106
|||||
201 TGGTTCTTCAATTGTGATGATGATGATGATGATGATGATGATGATGAT 250
106 AlaMetValAspSerAsnThrThrLeuGluAspAsnArgThrSerLys 122
|||||
251 TGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
123 AlaCysProGluHisSerAlaProIleLysValHisAsnGlnThrGly 139
|||||
301 GCGGTCCAGAGCATTCCTCCCATTAAGTTCATCATCATCAATCAAGCGG 350
139 YLysLysTyrGlnThrPaspAlaGlnThrGlnGlyTyrIleLeuGlySerP 156
|||||
351 TAAGAGTACCAATGGATGATGATGATGATGATGATGATGATGATGATGAT 400
156 hePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSer 172
|||||
401 TTTTATGCTGCTCATCATCATCATCATCATCATCATCATCATCATCATCAT 450
173 LysIleGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
|||||
451 AAAATAGGGGGGAAAATGCTGCTAGGATTTGGGATTCCTTGGCAGCTGCT 500
189 IleuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeu 206
|||||
501 CCTCACCTTTCACCTCCCTTTCGTCGACATTTAGAGATTGGACCTCACTCA 550
206 IeValLeuArgAlaLeuGlyLeuGlyLeuGlyValThrPheProAla 222
|||||
551 TTGTACACAGACACTAGAGAGACTAGAGAGAGGTTTACATTTCCAGCC 600
223 MetHisAlaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLe 239
|||||
601 ATGCATGCCATGTGCTCTTCTTGGGCTCCCTTGGGAAAGCAAGCAACT 650
239 uluSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
|||||
651 TCTTAGCATTTGCTATGACAGACAGCTTGGAGCGTAAATTTCTTTC 700
256 roleuSerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyr 272
|||||
701 CTCTTCTGGAATATTTGCTACTATATGATGATGATGATGATGATGATGAT 750
273 PhePheGlyThrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuVal 289
|||||
751 TTTTGGTACTATGGAATATTTGCTTCTTGGTCTTGGTCTTGGTCTTGGT 800
289 IserAspThrProGluHisLysArgIleSerHisTyrGlyLysGluTrp 306
|||||
801 TAGTGACACACCAAAAACCAAGAAATTTCCATTAAGAAAGGAAT 850
306 YIleLeuSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValPro 322

```

|||||
851 ACATCTTCATCATTAAGAATCAGCTTCTTCACAGAACTCAGTCCG 900
323 TrypValProIleLeuLysSerLeuProLeuTrpAlaIleValAlaIle 339
901 TGGGTACCATTTTAAATCCCTGCACCTTTGGCTATCGTAGTGCACA 950
339 sPheSerTrpAsnTrpPheTrpThrLeuLeuThrLeuLeuProThr 356
951 CTTTCTTACACAGCTGGACTTTTATCTTATTTACATTTATGCTTACTT 1000
356 YrMetLysGluIleLeuArgPheAsnValGlnGlnAsnGlyPheLeuSer 372
1001 ATATGAGAGAGATCCTTAAGTTCAATGTTCAAGAAATGGGTTTAACT 1050
373 SerLeuProTrpLeuGlySerTrpLeuLysMetIleLeuSerGlyGlnAl 389
1051 TCATTTGGCTTTTATGAGCTTTGGTTATGATGATCCCTGCTGTCAGAC 1100
389 AlaAspAsnLeuArgAlaLysTrpAsnPheSerThrLeuLysValArg 406
1101 TGCAGACATTTAAGCCAAATGGAATTTTCAACTTTATGCTTGCACA 1150
406 rGIIePheSerLeuIleGlyMetIleGlyProAlaValPheLeuValAla 422
1151 GAATTTTACCTTATAGATGATGATGACCTGCAGATTTCTGTGAGCT 1200
423 AlaGlyPheIleGlyCysAspTrpSerLeuAlaValAlaPheLeuThr 439
1201 GCTGGCTTCATGGCTGATGATTTCTTTGGCCGCTGCTTCCCTAACAT 1250
439 eSerThrThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHis 456
1251 ATCACAACAACATGGAGGCTTTTCTCTCTCGATTTAGCATCAACCATC 1300
456 euAspIleAlaProSerTrpAlaGlyIleLeuLeuGlyIleThrAsnThr 472
1301 TGGATATGCTGCTTCGATCTGATGATCCCTGCGCATTCACAAATGACA 1350
473 PheAlaThrIleProGlyMetValGlyProAlaIleAlaLysSerLeuTh 489
1351 TTTCCTCCTATTCAGAGATGATGGGCTCATGCTTAAATCTTGAC 1400
489 rProAspAsnThrValGlyLysTrpGlnThrValPheTrpIleAlaIle 506
1401 CCTGATTAACACCTGTGAGAAATGCAACCGCTTCTATATGCTGCTG 1450
506 IAlaIleAsnValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyLu 522
1451 CTATTAATGTTTGGTCCATTTTCTTACACTATTCGCCAAAGGTGAA 1500
523 ValGlnAsnTrpAlaLeuAsnAspHisHisGlyHisArgHis 536
1501 GTCAAAACCTGGCTCTCATGATCACCATGAGACACAGACAC 1542
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA158115
seq_documentation_block:
ID AA158115 standard: cDNA; 1488 BP.
XX
XX AA158115:
XX
XX 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 318.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX

```

```

OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX
XX P-PSDB; AAM38959.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 318; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 1488 BP; 368 A; 318 C; 323 G; 479 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 2614.00      Length: 495
XX      Ratio: 5.291      Gaps: 0
XX Percent Similarity: 99.798      Percent Identity: 99.798
XX
XX alignment_block:
XX US-09-776-865-2 x AA158115 ..
XX
XX Align seq 1/1 to: AA158115 from: 1 to: 1488
XX
XX 42 MetArgSerProValArgAspLeuAlaArgAsnAspGlyGlnGlnSerThr 58
XX |||||||
XX 1 ATGAGGCTCCGCGTTCGAGACATGCGCCGGAACGATGCGAGAGAGAC 50
XX
XX 58 rAspArgThrProLeuLeuProGlyAlaProArgAlaGlnAlaAlaPro 75
XX |||||||
XX 51 GGACCGGACCGCTCTTCTACCGGCGCGCCACGCGGGAAGCCGCTCCAG 100
XX |||||||
XX 75 aLysCysSerAlaArgTrpAsnLeuAlaIleLeuAlaPhePheGlyPhe 91
XX |||||||
XX 101 TGTGCTGCTGCTGCTTACACTTACAACTTATGCGCTTTTGGTTTC 150
XX

```

92 PheIleValIyralaLeuArgValAsnLeuSerValAlaLeuValaAspme 108
 151 TTCATTTGTATGCTATGCTGTAATCTGATGTTGCTTACGTGATAT 200
 108 tValAspSerAsnThrThrLeuGluuAspAsnArgThrSerLysAlaCysp 125
 201 GGTAGATTCAAAATACACTTAGAAGATTAATAGAACTTCCAGAGCGGTGC 250
 125 roGluHisSerAlaProIleLysValHisHisAsnGlnThrGlyLys 141
 251 CAGAGCATTTGCTCCCATAAAGTTTCATATCATATCAAGGGATAAAG 300
 142 TyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePhe 158
 301 TACCAATGGGATGCAGAACTCAAGAGATGATTCCTCGTTCTTTTATTA 350
 158 rGlyTyrllelleThrglnIleProGlyGlyTyrlValAlaSerLysIle 175
 351 TGGCTACATATCACACAGATTCCTGGAGATATGTTGCCAGCAAAATAG 400
 175 IyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValleuThr 191
 401 GGGGGAATAAGCTGCTGATGATTTGGGATCCTTGGCACTGCTGCTTACC 450
 192 LeuPheThrProIleAlaAlaAspLeuGlyValGlyProleuIleValle 208
 451 CTGTCTACTCCCATTCCTGCAGATTTAGAGATTGGACCACTCATTTTACT 500
 208 uArgAlaLeuGluGlyLeuGlyGlyValIleThrPheProAlaMetHis 225
 501 CAGAGCACTGAAGAGCTAGAGAGGGGTGTACATTTCCAGCCATGATG 550
 225 IamettPserSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 241
 551 CCATGTGCTTCTTGGGCTCCCTCTTGAAGAAACAACTTCTTACG 600
 242 ILeSerTyrlaGlyAlaGlnLeuGlyThrValIleSerLeuProLeuSe 258
 601 ATTTCAATATCAGAGACACAGCTTGGGACAGTAATTTCTTCTCTTCT 650
 258 rGlyIlelleCysTyrlMetAsnTrpThrTyrlValPheTyrlPheheg 275
 651 TGAATATATTTGCTACTATATGAAATGGACTTATGTCCTTCTACTTTT 700
 275 IyThrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerAsp 291
 701 GTACTATATGGAATATTTGGTTCTTTTGGATCTGGATTAAGTTAGTAC 750
 292 ThrProGlnLysHisLysArgIleSerHisTyrlGlyLysGluTyrlle 308
 751 ACACCAACAAACACAAAGAGATTTCCCATTTATGAAGAAAGATACATCT 800
 308 uSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValProTrpValP 325
 801 TTCATCATTTAAGAAATCAGCTTCTTCACAGAGTCCAGGCGCTGGGTAC 850
 325 roIleLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 341
 851 CCATTTTAAATCCCTGCCACTTTGGGCTATCTAGTATGACACTTTTCT 900
 342 TyrAsnTrpThrPheTyrlThrLeuLeuThrLeuLeuProThrTyrlMet 358
 901 TACAACCTGGACTTTTATCTTATGACATTTATGGCTCTACTATATGAA 950
 358 sGluIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeuP 375
 951 GGAATCTCAAGGCTCATGTTCAGAGAAAGGTTTATCTTCATATGC 1000
 375 roTyrlLeuGlySerTrpLeuCysMetIleLeuSerGlyGlnAlaAlaasp 391
 1001 CTATATTAGGCTCTGTGTATGATATCTCTGTCTGACAGCTGCTGAC 1050
 392 AsnLeuArgAlaLysTrpAsnPheSerThrLeuCysValaArgArgIleph 408

1051 AATTTAAGGCAAAATGCAATTTTCAACTTTATGTTGTCAGAAATTTT 1100
 408 eSerLeuIleGlyMetIleGlyProAlaValPheLeuValAlaIaGlyP 425
 1101 TAGCTTATAGAAATGATGATGAGCTGCAGATTTCTGTGAGCTGTGCT 1150
 425 heIleGlyCysAspTyrlSerSerLeuAlaValAlaPheLeuThrIleSerThr 441
 1151 TCATTTGCTGTGATATTTCTTTGGCGCTGTGCTTCCATACATATTCACA 1200
 442 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuaspI 458
 1201 ACATGGAGAGCTTTGCTCTCTCTGATTTAGCATCAACCATCGATAT 1250
 458 eAlaProSerTyrlAlaGlyIleLeuLeuGlyIleThrAsnThrPheAla 475
 1251 TGCTCTTCTGTATGCTGTGATCTCTCTGGGATCACAATATCATTTGCCA 1300
 475 hrIleProGlyMetValGlyProValIleAlaLysSerLeuThrProasp 491
 1301 CTATTCAGGAATGTTGGGCCGTCTATTTCTAAAGTCTGACCCCTGAT 1350
 492 AsnThrValGlyGluTrpGlnThrValPheTyrlleAlaAlaIleAs 508
 1351 AACACTGTGAGAAATGGCAAAACCGTGTCTATATTTCTGCTGATATTA 1400
 508 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGln 525
 1401 TGTTTTGGTCCATTTTCTTTTACACTATTTGCCAAAGGTGAAGTACAAA 1450
 525 snTrpAlaLeuAsnAspHisHisGlyHisArgHis 536
 1451 ACTGGGCTCTCATGATACACATGAGACAGACAC 1485
 seq_name: /SIDSL/gcgsdata/hold-geneseq/geneseg-emb1/NA2000.DAT:AAZ50876
 seq_documentation_block:
 ID AAZ50876 standard; cDNA; 2844 BP.
 AC AAZ50876;
 DT 31-MAY-2000 (first entry)
 DE Sheep GBS toxin receptor (SP55) cDNA.
 KW Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55;
 KW pathological vascularisation; cancer metastases; angiogenesis;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
 OS Ovis sp.
 XX
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 84..1571
 FT /tag= a
 FT /product= "Sheep GBS toxin receptor"
 PN WO200005375-A1.
 PD 03-FEB-2000.
 PF 22-JUL-1999; 99WO-US16676.
 PR 22-JUL-1998; 98US-0093843.
 PA (UYVA-) UNIV VANDERBILT.
 PI Hellerqvist CG, Fu C;
 DR WPI, 2000-205377/18.


```
506 aileasnvalpheglyalallephepetherleuphealalyglyluv 523
|||||
1478 TATCAATGATTTGGTCCATTTTCTTCACACTATTCGCCAAGGTGAAG 1527
523 aalgnasnrtpalaleuasnasphishisglyhisarghis 536
|||||
1528 TGCAGAACTGGCCATCAGTACACCAAGCACAGAAAC 1568

seq_name: /sids1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAD10326
seq_documentation_block:
ID AAD10326 standard; DNA; 2844 BP.
XX
AC AAD10326;
XX
DT 16-OCT-2001 (first entry)
XX
DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) DNA.
XX
KM Sheep: group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;
KM cytostatic; vulnerary; antiatherosclerotic; osteopathic; vasotropic;
KM prevention; attenuation; pathoangiogenic condition; cancer; scar;
KM wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KM keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KM vaccine; ds.
XX
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT CDS 84..1571
FT /tag= a
FT /product= "Sheep GBS toxin receptor protein (SP55)"
XX
PN WO200156598-A2.
XX
PD 09-AUG-2001.
XX
PF 02-FEB-2001; 2001WO-US03662.
XX
PR 02-FEB-2000; 2000US-0179870.
XX
PA (UYVA-) UNITV VANDERBILT.
XX
PI Helleqvist CG;
XX
DR WPI: 2001-488844/53.
DR P-PSDB: AAE06519.
XX
PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment -
XX
PS Disclosure; Page 47-50; 52pp; English.
XX
CC The present sequence is a DNA encoding sheep group B beta-haemolytic
CC Streptococci (GBS) toxin receptor protein, SP55. The present invention
CC relates to a method for preventing or attenuating a patho-angiogenic
CC condition in a mammal which comprises administering to the mammal one
CC or more GBS toxin receptors or their immunogenic fragments to induce
CC or maintain an immune response to one of GBS toxin receptors. The
CC method is useful for preventing or ameliorating pathoangiogenic
CC conditions such as cancer, scarring during wound healing, gliosis
CC during repair of nerve injury, chronic wounds, keloids, reperfusion
CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
CC psoriasis in mammals. The proteins of the invention are also used
CC as vaccines.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;
```

alignment_scores:

```
Quality: 2338.00 Length: 497
Ratio: 4.901 Gaps: 0
Percent Similarity: 95.976 Percent Identity: 86.117

alignment_block:
US-09-776-865-2 x AAD10326 ..
Align seg 1/1 to: AAD10326 from: 1 to: 2844

40 GlyValMetArgSerProValArgAspLeuAlaArgAsnAspGlyIugl 56
|||||
78 GGCATCATGTAAGTCCCGGTTTCGACCTTAGCCCGGACGCGGCGAGGA 127
56 userThrAspArgThrProleuLeuProGlyAlaProArgAlaGluAla 73
|||||
128 GGGCTCGGACGCGCACACCGCTCCTGCAGCGCGCCGCGGGCGGAACCCG 177
73 LaProValCysCysSerAlaArgTyrAsnLeuAlaIleuAlaIlePhe 89
|||||
178 CTCACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
90 GlyPhePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuVa 106
|||||
228 GGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 277
106 LasPheValAspSerAsnThrThrLeuGluAspAsnArgThrSerLysa 123
|||||
278 GGACATGCTGATTCACAAACAACTGCCAAGATATAGAACCTCCACAG 327
123 LacGlyProGluHisSerAlaProIleLysValHisAsnGlnThrGly 139
|||||
328 AGTGTGCGAGCATTCCTGCTCCATTAAGTCTTCACACCAACGCGGT 377
140 LysLysTyrGlnThrAspAlaGluThrGlnGlyTyrIleLeuGlySerPh 156
|||||
378 AAAAGTACCGGGTGGATGCAAACTCAAGATGATTCGCGATCTTT 427
156 ePheTyrGlyTyrIleIleThrGlnIleProGlyLysTyrValAlaSerL 173
|||||
428 TTTCTATGCTACATCATCACACAATTCCTGAGAGATATGTCGCCAGCA 477
173 LysIleGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
|||||
478 GAAGTGGGGGGAAGCTGCTGAGATTCGGATTCGCTGCTGCTGCTGCTGCT 527
190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyAlaGlyProLeuI 206
|||||
528 TTCACCTGTTCACTCCCTCGCTGCGAGATTCGAGTCCGACCCCTTGT 577
206 eValLeuArgAlaLeuGluIuglLeuGlyIuglIuglValThrProAlaM 223
|||||
578 TGCATCAGGCGCACTAGAAGGCTAGAGAGGCTGCATCATTCACAGCA 627
223 ethAlaMetTrpSerSerTrpAlaProProleuGluArgSerLysLeu 239
|||||
628 TGCATGCCATGCTGCTTCATGGGCTCCCTCTTGAAGAACAGACCTT 677
240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
|||||
678 CTGAGTATTTTCATATGAGAGGACCAACTGGAGCACTGATTTCTCTCC 727
256 oluSerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValAlaPheTyr 273
|||||
728 TCTTTCGAGATATTTGCTACATATAGATTTGACTATGCTTCTGAT 777
273 hebPheGlyThrIleGlyIlePheThrPheLeuLeuThrIleTrpLeuVal 289
|||||
778 TCTTTCGATTTGTTGAATCATCTGGTTATTTATAGATCTCTTAGT 827
290 SerAspThrProGlnLysHisLysArgIleSerHisTyrGluLysGluT 306
|||||
828 AGTATACACCAAGAACTCACAAGACATTCCTCCATGATAAAGAGATA 877
```



```

FT FT /transl_except= (pos:1270..1272; aa:Xaa)
FT FT /transl_except= (pos:1291..1293; aa:Xaa)
FT FT /transl_except= (pos:1306..1308; aa:Xaa)
FT FT /transl_except= (pos:1315..1317; aa:Xaa)
FT FT /transl_except= (pos:1324..1326; aa:Xaa)
FT FT /transl_except= (pos:1330..1332; aa:Xaa)
FT FT /transl_except= (pos:1333..1335; aa:Xaa)
FT FT /transl_except= (pos:1339..1341; aa:Xaa)
FT FT /transl_except= (pos:1348..1350; aa:Xaa)
FT FT /transl_except= (pos:1357..1359; aa:Xaa)
FT FT /transl_except= (pos:1372..1374; aa:Xaa)
FT FT /transl_except= (pos:1375..1377; aa:Xaa)
FT FT /transl_except= (pos:1381..1383; aa:Xaa)
FT FT /transl_except= (pos:1386..1388; aa:Xaa)
FT FT /transl_except= (pos:1396..1398; aa:Xaa)
FT FT /transl_except= (pos:1402..1404; aa:Xaa)
FT FT /transl_except= (pos:1420..1422; aa:Xaa)
FT FT /transl_except= (pos:1444..1446; aa:Xaa)
FT FT /transl_except= (pos:1456..1458; aa:Xaa)
FT FT /transl_except= (pos:1459..1461; aa:Xaa)
FT FT /transl_except= (pos:1462..1464; aa:Xaa)
FT FT /transl_except= (pos:1471..1473; aa:Xaa)
FT FT /transl_except= (pos:1483..1485; aa:Xaa)
FT FT /note= "Xaa = unknown"

PN PN WO200005375-A1.
PN PN 03-FEB-2000.
XX XX
XX XX 22-JUL-1999; 99WO-US16676.
XX XX
XX XX 22-JUL-1998; 98US-0093843.
XX XX
XX XX (UYVA-) UNIV VANDERBILT.
XX XX
XX XX Hellegqvist CG, Fu C;
XX XX
XX XX WPI: 2000-205377/18.
XX XX P-PSDB: AAY45090.

Alignment_scores:
Quality: 2217.00 Length: 494
Ratio: 5.073 Gaps: 0
Percent Similarity: 88.462 Percent Identity: 86.437

alignment_block:
US-09-776-865-2 x AAZ50880 ..

Align seg 1/1 to: AAZ50880 from: 1 to: 1485

42 MetAaSerProValArgAspLeuAlaArgAsnAspGlyGluSerTh 58
||||| ||||||| ||||||| ||||||| |||||||
1 ATGAGCTGTCGCGGTTTSGACCTRGCCCGSARCAAGCGCGAGRCMC 50

58 rAspArgThrProIleuLeuProGlyAlaProArgAlaGluAlaProV 75
||||| ||||||| ||||||| ||||||| |||||||
51 GGACCGCACRCKCTCTCRMGSGCGCCCGGCGGCGSAGSCGCTCAG 100

75 aLcysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGly 91
||||| ||||||| ||||||| ||||||| |||||||
101 TRTGCGTCTGCGTCGTTACAACTAGACAMTTTTGKCCCTTTTGGTTT 150

92 PheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAsp 108
||||| ||||||| ||||||| ||||||| |||||||
151 TTCRTTSTSTTKCATTAACGKTGAATCTGAGYGTGCTGTAGGGAAT 200

108 tValAspSerAsnThrThrLeuGluAspAsnArgThrSerTyrAlaCys 125
||||| ||||||| ||||||| ||||||| |||||||
201 GGTGATTTCAATATCAACTKTMRAGATTAATAGAACTTCWASGMSGT 250

125 rGluHisSerAlaProIleValValHisHisAsnGlnTrnTrnGlyVal 141
||||| ||||||| ||||||| ||||||| |||||||

```


FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..199; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..252; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)
FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)
FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)
FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)

FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1081..1083; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)
FT /transl_except- (pos:1348..1350; aa:Xaa)
FT /transl_except- (pos:1357..1359; aa:Xaa)
FT /transl_except- (pos:1372..1374; aa:Xaa)
FT /transl_except- (pos:1375..1377; aa:Xaa)
FT /transl_except- (pos:1381..1383; aa:Xaa)
FT /transl_except- (pos:1384..1386; aa:Xaa)
FT /transl_except- (pos:1396..1398; aa:Xaa)
FT /transl_except- (pos:1402..1404; aa:Xaa)
FT /transl_except- (pos:1420..1422; aa:Xaa)
FT /transl_except- (pos:1444..1446; aa:Xaa)
FT /transl_except- (pos:1456..1458; aa:Xaa)
FT /transl_except- (pos:1459..1461; aa:Xaa)
FT /transl_except- (pos:1462..1464; aa:Xaa)
FT /transl_except- (pos:1471..1473; aa:Xaa)
FT /transl_except- (pos:1483..1485; aa:Xaa)
/note- "Xaa = unknown"

MO200005375-A1.

XX 03-FEB-2000.

XX 22-JUL-1999; 99MO-US16676.

XX

[illegible][illegible]

```

XX      20-NOV-2001 (first entry)
DT
XX      Human sodium dependent phosphate co-transfer protein 35 cDNA.
DE
XX      Human; sodium dependent phosphate co-transfer protein 35;
KW      hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW      nephritis; gene therapy; ss.
XX
XX      Homo sapiens.
XX
XX      CN1298882-A.
XX
XX      13-JUN-2001.
PD
XX      06-DEC-1999; 99CN-0124217.
PE
XX      06-DEC-1999; 99CN-0124217.
PR
XX      06-DEC-1999; 99CN-0124217.
XX
XX      (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
XX      Mao Y, Xie Y;
XX
XX      WPI: 2001-503367/56.
DR      P-PSDB; AAG65238.
XX
XX      Human Na-dependent phosphate cotransporter 35 and its coding sequence -
PT
XX
XX      Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
XX
XX      The present invention provides the protein and coding sequences of human
XX      sodium dependent phosphate co-transfer protein 35. The sequences can be
XX      used in the treatment of hypophosphaturia, hypercalcaemia,
XX      hypophosphataemic rickets and nephritis. The present sequence is the
XX      coding sequence of the invention.
XX
XX      Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;
SQ

```

alignment_scores:

Quality:	1716.00	Length:	321
Ratio:	5.346	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-776-865-2 x AAH79234 ..

Align seg 1/1 to: AAH79234 from: 1 to: 2670

```

216 GIUGLYVALThrPheProAlaMetHisAlaMetTTPSerSerTIPAlaPr 232
|||||
3 GAGGGGTTCATTTCCAGCCATGCGATGCGATGCTCTTGGGCTCC 52
232 oProLeuGluArgSerLysLeuSerLysSerLysSerLysLysLysLysL 249
|||||
53 CCCTCTTGAAAGAGAGAACTCTTAGCATTCATATGACAGAGACAGC 102
|||||
249 euGLYThyValIleSerLeuProLeuSerGlyIleIleCysTyrTrpMet 265
|||||
103 TTGGGACAGTAATTTCTCTCTCTTCTGGAATTAATTTGCTACTATATG 152
|||||
266 AsnTrpThrValPheTyrPhePheGlyTrpIleGlyIlePheTrpH 282
|||||
153 AATTGGACTTATGCTCTTCTACTTTTGTGCTATTTGGAATATTTGGTT 202
|||||
282 eLeuLeuTrpIleTyrPheValSerAspThrProGlnLysHisLysArgI 299
|||||
203 TCTTTTGTGATCTGGTAGTAGTACACACACACACACACACACACAGAGA 252
|||||
299 lSerHisTyrGlnLysGlnLysGlnLysSerSerLeuArgAsnGlnLeu 315
|||||
253 TTTCCATTATGATAAAGAAATCATCTTTTCATCATTAAGAATACAGCTT 302

```

```

316 SerSerGlnLysSerValProTyrValProIleLeuLysSerLeuPro 332
|||||
303 TCTTCACAGAACTCAGTCCGCTGGTACCATTTTAAATCCCTGGCACT 352
|||||
332 uTPAlaIleValAlaHisPheSerTyrAsnTrpThrPheTyrThrL 349
|||||
353 TTGGGCTATGCTAGTGTGACACTTCTTCAACAAGTGGACTTTTATACCT 402
|||||
349 euleuThrLeuLeuProThrTyrMetLysGlnIleLeuArgPheAsnVal 365
|||||
403 TATTGACATTATTTGCTCTACTTATATGAGAGAGATCTTAAGTTCAAVGT 452
|||||
366 GlnGluAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeu 382
|||||
453 CAAGAGAAATGGGTTTATCTTCATTCATTCCTTATTTAGGCTCTGGTTAG 502
|||||
382 smetIleuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsn 399
|||||
503 TATGATCCTGCTGTCGCAAGCTGCTGACAAATTTAAAGGCCAAATGGAAT 552
|||||
399 heserThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGly 415
|||||
553 TTTCAACTTATGCTTCCGAGAAATTTTAAAGCTTATAGGAATGATTTGA 602
|||||
416 ProAlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyrSer 432
|||||
603 CCTGCAGTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 652
|||||
432 uAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSers 449
|||||
653 GGCCGTTGCTTCTTAACATATATCAACACACTGGAGAGCTTTGCTGCTT 702
|||||
449 erGLYPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIle 465
|||||
703 CTGGAATTAAGCATCAACCAATCGATATTTGCTCTGATCTGCTGATATC 752
|||||
466 LeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyPr 482
|||||
753 CTCCTGGGCAATCAAAATACATTTGCCACTATTTCCAGGAATGCTGGGCC 802
|||||
482 oValIleAlaLysSerLeuThrProAspAsnThrValGlyIleTyrGlnT 499
|||||
803 CGTCATTTGCTAAAGCTGACCCCTGATACACACTGTTGGAGATGGCANA 852
|||||
499 hrValPheTyrIleAlaAlaIleAsnValPheGlyAlaIlePhePhe 515
|||||
853 CCGTGTCTTATATGCTGCTGCTATTAATGTTTGTGCTCCATTTTCTTT 902
|||||
516 ThrLeuPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisH 532
|||||
903 ACACATATTTGCCAAGAGTAACTACAAAAGCTGGCTCTCAATGATCACCA 952
|||||
532 sGLYHisArgHis 536
|||||
953 TCGACACACAGACAC 965

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI59901

seq_documentation_block:

ID	AAI59901	standard;	cdna;	929	BP.
XX	AAI59901;				
AC	22-OCT-2001	(first entry)			
DT	Human polynucleotide SEQ ID NO 3890.				
XX					
DE	Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;				
XX	peripheral nervous system; neuropathy; central nervous system; CNS;				
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;				
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;				
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;				
KW	leukemia; ss.				

XX Homo sapiens.
OS
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB: AAM40745.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3890; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 929 BP; 214 A; 224 C; 223 G; 268 T; 0 other;

alignment_scores:
Quality: 1614.00 Length: 304
Ratio: 5.309 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-776-865-2 x AA159901 ..

Align seg 1/1 to: AA159901 from: 1 to: 929

8 ProProArgProValGlnProAlaArgProGlyGlyPheGlyLeuSerG1 24
17 CCGCGCCGCCGCGGCGCAGCGCTGCGGCGGCGCTTGGCGCTGTGGG 66
24 YArgGArgSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGly 41
67 CCGGCGCTCCCTTCTCTGCGCAGGTGGCGAGTACACCTGCTCAGTAGCG 116
41 AlMetArgSerProValArgAspLeuAlaArgAsnAspGlyGluGluSer 57
117 TCTATGAGTCTCGGGTTCGAGACCTGGCCGGAACGATGGCGAGAGAGC 166

58 ThrAspArgThrProLeuLeuProGlyValAlaProArgAlaGluAlaApr 74
167 ACGGACCGCAGCGCCTCTTCTACCGGGCGCCCGACGGCGCGACGCGCTCC 216
74 OValCysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyP 91
217 AGTGTGCTGCTCTGCTGTTACCACTTAGCAATTTGGCTTTTGTGTT 266
91 hePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAsp 107
267 TCTTCATTTGTATGATGCAATGCTGATCGATCGATGCTGGCTAGTGAT 316
108 MetValAspSerAsnThrThrLeuGluAspAsnArgThrSerTyrAlaCy 124
317 ATGCTACATTCATCAATACACTTTAGAAAGATATAGAACTTCCAAAGCGCTG 366
124 sProGluHisSerAlaProIleLysValHisHisAsnGlnThrGlyLysL 141
367 TCCAGACCATTCCTGCTCCATMAAAGTTCATCAATCAAAACGGGTAGA 416
141 YsTyrGlnTyrPaspAlaGluThrGlnGlyThrIleLeuGlySerPhePhe 157
417 AGTACCAATGGATGAGAACTCAAGATGATTCGCGTTCCTTTT 466
158 TyrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerLysI 174
467 TATGGCTACATCATCACACAGATTCCTGGAGGATATGTCGACGACCAAT 516
174 eGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeuT 191
517 AGGGGGGAAATGCTGCTAGATTGGGATCTYGGGACGCTGCTCTCA 566
191 hrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleVal 207
567 CCTGTTCACTCCCATTTGCTGCGAGATTTAGGAGTTGGACCACTCATTTGTA 616
208 LeuArgAlaLeuGlyGlyLeuGlyGlyValThrPheProAlaMetThr 224
617 CTGAGACACTAGAGAGAGTAGAGAGGAGGTTACATTTCCACAGCANTGA 666
224 sAlaMetThrSerSerThrPalAProProLeuGluArgSerLysLeuLeuS 241
667 TGGCATGTGCTCTTGGGCTCCCTCTTGAAGAACCAAACTCTCTTA 716
241 eIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuProLeu 257
717 GCATTTGATATGACAGAGACAGCTTGGACAGTAATTTCTTCCCTCTT 766
258 SerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPheP 274
767 TCTGGAATATATTGCTACTATGATGAATGGACTTATGCTCTTACTTTT 816
274 eGlyThrIleGlyIlePheThrPheLeuLeuTrpIleTrpLeuValSerA 291
817 TGGTACTATTGGAATATTTTGGCTTCTTTTCTGATCTGCTGTAGTACTG 866
291 sPThrProGlnLysHisLysArgIleSerHisTyrGlnLysGluTyrIle 307
867 ACACACCACAAAAACACAGAGAGATTTCCCATTTATGAAAAAGAAATACATT 916
308 LeuSerSerLeu 311
917 CTTTCAATCATTA 928
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseg-emb1/NA2001A.DAT:AAK93901
seq_documentation_block:
ID AAK93901 standard; cDNA; 853 BP.
XX
AC AAK93901;
XX
DT 06-NOV-2001 (first entry)

XX 07-JUL-2000; 2000EP-0114089.
 PF 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX (HELI-) HELIX RES INST.
 PI Ora T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 2; SEQ ID NO 824; 1380pp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 5'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SO Sequence 838 BP; 189 A; 196 C; 210 G; 237 T; 6 other;

alignment_scores:
 Quality: 1339.50 Length: 281
 Ratio: 4.998 Gaps: 2
 Percent Similarity: 95.374 Percent Identity: 94.306

alignment_block:
 US-09-776-865-2 x AAK92364 ..

Align seg 1/1 to: AAK92364 from: 1 to: 838

```

15 AAlaArgProGlyGlyPheGlyLeuSerGlyArgArgSerLeuLeuLeuGly 31
   |||||||
2  GCTCGGCGCGGCGGCTTGGGCTGCGGCGCGGCTCGCTCTCTGCGCA 51
   |||||||
31 nValAlaSerThrProAlaHisValGlyValMetArgSerProValArg 48
   |||||||
52 GGTGCGAGTACACCTGCTCACTGAGGCGTCAAGGCTCGGTTGCGAG 101
   |||||||
48 sPLeuAlaArgAsnAspGlyGluSerThrAspArgThrProLeuLeu 64
   |||||||
102 ACCTGCGCCCGAAGCATGCGAGGAGACACGACCGACGCTCTCTCTTA 151
   |||||||
65 ProGlyAlaProArgAlaGluAlaAlaProValCysCysSerAlaArgTy 81
   |||||||
152 CCGGGGCGCCCGGCGGCGGAGCCGCTCAGTGTCTCTGCTGCTGCTTA 201
   |||||||
81 rAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeu 98
   |||||||
202 CAACCTTACCAATTTGGCCTTTTGGTTTCTCTCATTTGTATGATTAAC 251
   |||||||
98 rGValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThr 114
   |||||||
252 GTGTGAATCTGAGTGTGCTGTTAGTGAATATGATGATTCAAATCAACT 301
   |||||||
115 LeuGluAspAsnArgThrSerTyrAlaCysProGluHisSerAlaProI 131
   |||||||
302 TTAGAAGATTAATAGAACTTCCAGGCGTGTCCAGACATTCCTGCTCCAT 351

```

```

131 eLysValHisHisAsnGlnThrGlyLysTyrGlnTrpAspAlaGluTr 148
   |||||||
352 AAAAGTTCATCATATATCAAAACGGGTAAAGACGTACCAATGGATGCAAAA 401
   |||||||
148 hGlnGlyTrpIleLeuGlySerPhePheTyrGlyTyrIleIleThrGln 164
   |||||||
402 CTCAGAGATGAGATCTCGGTTCTTTTATGCTATCATCATCAGACAG 451
   |||||||
165 IleProGlyGlyTyrValAlaSerTyrIleGlyGlyLysMetLeuLeu 181
   |||||||
452 ATTCCTGAGCATATGTTGCCAGCAAAATAGGGGGGAAAATGCTGCTGAG 501
   |||||||
181 yPheGlyIleLeuGlyThrAlaValLeuThrLeuPheThrProIleAla 198
   |||||||
502 ATTTGGATTCCTGGGACCTGCTGCTCACCCTGCTTCCATCCCATGCTG 551
   |||||||
198 lAspLeuGlyValGlyProLeuIleValLeuArgAlaLeuGluGlyLeu 214
   |||||||
552 CAGATTTTAGAGTTGAGCACTCATGTCACAGACCTAGAAAGACTA 601
   |||||||
215 GlyGluGlyValThrPheProAlaMetHisAlaMetTrpSerSerTrpAl 231
   |||||||
602 NGAGAGGCTGTATCACTTTCCAGCCATGCATGCATGTGCTTCTTGCGC 651
   |||||||
231 aProProLeuGluArgSerLysLeuLeuSerLysSerTyrAlaGlyAlaG 248
   |||||||
652 TCCTCCCTTGGAAGAACCAACTCTTAGCATTTTCATATGACAGAGCAC 701
   |||||||
248 lLeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyr 264
   |||||||
702 ANCTTGGGACAGTAATTTCTCTTCTCTGCAATTAAT.TGCTACTAT 750
   |||||||
265 MetAsnTrpThrTyrValPheTyrPhePheGlyThrIleGlyIlePheTr 281
   |||||||
751 ATGATTTGGGACTTATGCTTCTACTTTTGGGACTATATGGAATATTGG 800
   |||||||
281 pPheLeuLeuTrp.IleTrpLeuValSerAspThrProGln 294
   |||
801 GTTC...TTTGGGAGCTGCTGTAANTANTGACACCCCAAA 838
   |||

```

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn_emb1/NA2001B.DAT:ABU03769

seq_documentation_block:
 ID ABL03769 standard; cDNA; 1939 BP.
 XX
 AC ABL03769;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5789.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-0509231.
 XX
 PF 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 DR WPI; 2001-656860/75.
 DR P-PDB; ABB59666.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more

us-09-776-865-2.p2n.png

214 euglygluglyvalthrpheprolamethisalamettrpsersertrp 230

879 TGGCGAGGCGTCACCTATCCAGCTATGCACGCCATGCTTGCCCACTGG 928

231	AlaProProlLeuGluArgSerLysLeuLeuSerLleSerTyrAlaGlyAl	247
929	ATACGCGCCCTGGAGAGGAAACAAGTCCGCCCAACTCTATCGCGGCTC	978
247	aglnleuGlyThrValIleSerLeuProlSerGlyTleLeu...	266
979	CAATATCGGAACAGCATTTCCATGCCGCTGGCGGATGGCTGGTCGC	1020
263TyrTylMetasnrPrpThrTyrValPheTyrPhepneGlyThrIle	277
029	TGGACTTCTGTGGGCGTGGACCGCTGCGCTTTCTCAATCTTTCGGAGCTGTG	107

278 GlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerAspThrProG1 294

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

129 CGATCATCCAGGATCTCAGAATCGAGCGAGAGTATATCGAAAGGAGTTC 117

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible][illegible][illegible]

339 | eDheSarturAentromhrrDbhottumhri ouI oumhri ouf ouDromhrr 355

329 AATGCGGCGAGGCGATGCGCGCTTCTTACAGCGCGCTTAAAGGAGCTGCGCGACCG 1337

356 TyrMetIsvG]uT]eLeuArpAAsnVa]G]nG]uAsnG]vpheI.euSe 372

379 TACATGAGCAACATCTACACTTTGACATCCAGTCGAATGCTCTGCTCAA 142

372 rSerLeuProTyrLeuGlySerTrpLeuCysMetIleLeuSerGlyGlnA 389

429 TCGGCTGCCGTACTAACCTCCTGTTGTTGGGCAATGGCTGCTCGCCC 147

389 1aAlaAspAsnLeuArgAlaIaIysTrpAsnPheSerThrLeuCysValArg 405

479 TGGCGATTGGATGCTAGCCAGACGCTACATATCGCTGCTGAACCTCGTAT 152

406 ArgIlePheSerLeuIleGlyMetIleGlyProAlaValPheLeuValAl 422

529 AAGTTGTGGAACACGCTGGCCTCCGTGGTGCCATCACTGGGCCGTGATTGG 157

422 aaLaGlyPheIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrI 439

579 CATCATCTATGTGGGCTGGGATTTGGGTGTGGGTCACCTTTATGTGGCCG 162

[illegible]

0027 9C6A9B6C1C111C9BC9CC6C1C1M1B1G9CAACCGAMIGMAICAC 187

```

130 pcnnspleneiafiocei y i n d u i y i i e d e u d e u i y i e i m i n s m i n 4 / 2
    :: ::::: ||| ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    :: ::::: ||| ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

OM of: US-09-776-865-2 to: EST:* out_format : pfs

Date: Jul 15, 2002 2:32 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed.p2n.model -DEV=xlh
-O=/cgn2.1/USPPO.spool/US09776865/runat.15072002.082207.27382/app-query.fasta.1.1159
-DB=BSF -QMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPEXT=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=200 -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09776865@cgn1.1.3468
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-776-865-2

Database: EST*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 2434.030000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:AL550137	1107.00	2746.90	9.3e-144	985	AL550137 LTI_NFL006.P1
gb_est2:BI907284	1136.00	1986.56	2.1e-101	754	BI907284 60306385881 NIH_MGC_11
gb_est2:BI661062	1115.00	1789.83	1.9e-90	801	BI661062 60330436281 NIH_MGC_11
gb_est2:BI868761	1062.00	1705.79	9.1e-86	676	BI868761 60144312761 NIH_MGC_65
gb_est2:BI868052	1024.50	1644.09	2.5e-82	736	BI868052 60338678761 NIH_MGC_87
gb_est2:BI676755	991.00	1591.57	2.1e-79	626	BI676755 60208438061 NIH_MGC_83
gb_est2:BI697765	991.00	1590.18	2.1e-79	711	BI697765 60334858061 NIH_MGC_83
gb_est2:BI597146	930.50	1489.64	1.0e-73	885	BI597146 60205500961 NCI_CGAP_5
gb_est2:BI141137	909.00	1454.69	8.8e-72	904	BI141137 60176083861 NCI_CGAP_5
gb_hic:BC023331	909.00	1448.95	1.8e-71	1531	BC023331 Mus musculus, clone 1
gb_est2:BI86541099	900.00	1440.24	5.6e-71	833	BI86541099 60257026561 NIH_MGC_77
gb_est1:BI8613552	893.00	1441.55	4.8e-71	682	BI8613552 BI8613552 RIKEN full-16
gb_est1:BI8537525	863.50	1384.64	7.1e-68	645	BI8537525 BI8537525 RIKEN full-16
gb_est1:AL597124	862.00	1385.00	6.7e-68	499	AL597124 DKFZP313H1911_r1 313
gb_est1:BA833287	802.00	1287.36	1.8e-62	518	BA833287 uc050609.r1 Soares.NHM
gb_est2:BI8689819	793.00	1269.50	1.8e-61	700	BI8689819 6014465161 NIH_MGC_65
gb_est2:BI8686603	782.00	1246.75	3.4e-60	1100	BI8686603 60278557761 NCI_CGAP
gb_est2:BI817031	777.50	1245.48	4.0e-60	633	BI817031 UMN07804 Canine Brain
gb_est1:BB610013	772.50	1243.10	5.4e-60	630	BB610013 RIKEN full-16
gb_est1:BB644731	772.50	1237.44	1.1e-59	629	BB644731 RIKEN full-16
gb_est2:BI851890	711.50	1138.10	3.8e-54	658	BI851890 60337804761 NCI_CGAP_5
gb_est2:BM179717	707.50	1132.86	7.4e-54	587	BM179717 da111810.y1 NICHD XGC
gb_est2:BU060500	682.00	1089.30	2.0e-51	721	BU060500 BU060500 NIBB Mochi1
gb_est1:AA258513	681.00	1093.26	1.2e-51	432	AA258513 z559001.r1 Soares.NHM
gb_est2:BI563945	662.00	1062.55	6.1e-50	429	BI563945 UT-R-C4-akw-c-03-0-UT
gb_est2:BU031727	639.00	1021.06	1.3e-47	632	BU031727 BU031727 NIBB Mochi1
gb_est2:BI445533	539.00	858.75	1.4e-38	647	BI445533 dae81109.y3 NICHD XGC
gb_est2:BI459662	532.50	845.76	7.3e-38	811	BI459662 AT14282.Sprtime AT Dros
gb_est2:BI375421	530.00	848.25	5.3e-38	445	BI375421 60123048561 NCI_CGAP_5
gb_est2:BI63685	520.50	834.95	2.9e-37	367	BI63685 y55601.r1 Soares.fetal
gb_est2:BI873013	517.00	819.93	2.0e-36	866	BI873013 60339819361 NIH_MGC_94
gb_est2:BI123545	503.50	799.30	2.8e-35	772	BI123545 6017588661 NCI_CGAP_5
gb_est2:BI892556	499.50	792.99	6.4e-35	760	BI892556 6014333581 NIH_MGC_72
gb_est2:BI13154	494.00	794.16	5.5e-35	301	BI13154 yx53h06.r1 Soares.melan
gb_est2:BI734022	482.00	768.79	1.4e-33	753	BI734022 6033555761 NIH_MGC_94
gb_gss:CMS0069V	482.00	767.24	6.2e-33	1075	AL053164 Drosophila melanogast
gb_est2:BI753164	476.50	757.24	6.2e-33	661	BI753164 Tetracodon nigroviridis
gb_gss:CMS0352N	472.00	746.08	2.6e-32	943	AL223388 Tetracodon nigroviridis
gb_est2:BF45471	461.50	732.64	1.5e-31	679	BF45471 AT04410.Sprtime AT Dros
gb_est2:BU106525	439.50	697.10	1.4e-29	672	BU106525 unpublished

seq_name: gb_est1:AL550137

seq_documentation block:

LOCUS AL550137 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.P12 Homo sapiens cDNA clone CS001040YK17 5
prime, mRNA sequence.

ACCESSION AL550137
VERSION AL550137.1 GI:12886813
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 985)
Li W.B., Gruber C., Jessee J. and Polayes D.
Full-length cDNA libraries and normalization
Unpublished (2001)

COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..985
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS001040YK17"
/clone_id="LTI_NFL006.P12"

/tissue="type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 236 a 224 c 224 g 300 t 1 others

ORIGIN

alignment_scores: Quality: 1707.00 Length: 328
Ratio: 5.220 Gaps: 0
Percent Similarity: 99.695 Percent Identity: 99.390

alignment_block: US-09-776-865-2 x AL550137 ..

Align seg 1/1 to: AL550137 from: 1 to: 985

23 SerclyARGARGSerLeuLeuGysGlnValAlaSerThProAlaHisVa 39

3 TCGGCGCGGCGCGCTTCCTTCGAGGCGGCGGCGGCGGCGGCGGCGG 52

39 IGIVAlMeIArgSerProValArgSplLeuAlaArgAsnAGIYGLUG 56

53 AGCGGTCATGAGGTCTCCGTTTCGAGACCTGCGCGGAGAGGCGGAGC 102

56 IuseThrsparqThrProLeuLeuProGlyAlaProArgAlaGluAla 72

103 AGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 152

73 AlaprovAlCysCysSerAlaArgTYraSnLeuAlaIleLeuAlaPhep 89

153 GCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 202

```

89 eglYpHePheIleValTYrAlaLeuArgValAsnLeuSerValAlaLeu 106
|||||
203 TGGTTTCTTCATGTGTGATGCATTACGTGTGATCTGATCTGGCTTGA 252
|||||
106 AlaSPMeValAspSerAsnThrThrLeuGluAspAsnArgThSerLys 122
|||||
253 TGGATATGTGTAGATTCAAAATACAACTTTAGAAAGATAATAGAACTTCCA 302
|||||
123 AlaCysProGluHisSerAlaProIleLeuValHisHisAsnGlnThrG 139
|||||
303 GCGTCCAGAGCATCTGCTCCCAATAAAGTTCACTAATACAAACGGG 352
|||||
139 YLysLYSTyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySer 156
|||||
353 TAAGAAGTACCAATGGATGCAGAAACTCAAGAGATGATCTCGGTCTCT 402
|||||
156 hePheTYrGlyTYrIleIleThrGlnIleProGlyGlyTYrValAlaSer 172
|||||
403 TTTTATAGGCTACATCATCACACAGATTCTCGAGATATGTTGCCAGC 452
|||||
173 LysIleGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVa 189
|||||
453 AAATA.GGGGGGAAAATGCTGCTAGGATTGGGATCTTGCCACTGCTGT 501
|||||
189 IleuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeu 206
|||||
502 CCTACCCCTGTTCACTCCCATTTGGCGAGATTTGGAGATTGGACACATCA 551
|||||
206 leValLeuArgAlaLeuGluGlyLeuGlyGlyValThrPheProAla 222
|||||
552 TTGACTAGAGACACTAGAAAGACTAGAGAGGGGTTCACATTTCCACCC 601
|||||
223 MethIAsaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLe 239
|||||
602 ATGCATGCCATGTGCTCTTCTTGGGCTCCCTCTTGAAAAGACCAACT 651
|||||
239 uLeuSerIleSerTYrAlaGlyAlaGlnLeuGlyThrValIleSerLeu 256
|||||
652 TCTTAGCATTTCAATATGCAGAGACACAGCTTGGGACAGTAATTTCTTTC 701
|||||
256 roLeuSerGlyIleIleCysTYrTYrMetAsnTrpThrTYrValPheTYr 272
|||||
702 CTCCTTCTGGAAATATTTGCTACTATAGAAATGGAACTTATGCTCTTAC 751
|||||
273 PhePheGlyThrIleGlyIlePheThrPheLeuLeuTrpIleTrpLeuVa 289
|||||
752 TTTTGTGTTACTATTTGAAATATTTGTTCTTTTGTGTGATCTGTTAGT 801
|||||
289 lSerAspThrProGlnLysHisLysArgIleSerHisTYrGlnLysGln 306
|||||
802 TAGTGACACACACAAAAACACAAGAGAAATTTCCCATATGAAAGGAAT 851
|||||
306 YrIleLeuSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValPro 322
|||||
852 ACATCTTTCATTCATTAAGAAATCAGCTTCTTCCACAGAACTCAGTGGCG 901
|||||
323 TrpValProIleLeuLysSerLeuProLeuTrpAlaIleValAlaHis 339
|||||
902 TGGGTACCCATTTTAAATCCCTGCCACTTTGGGCTATCGTAGTTGCACA 951
|||||
339 sPheSerTYrAsnTrpThrPheTYrThrLeuLeu 350
|||||
952 CTATTCTTACAACTGRCCTTTTATACTTATTTG 985
|||||
seq_name: gb_est2:BI907284
seq_documentation_block:
LOCUS BI907284 754 bp mRNA linear EST 16-OCT-2001
DEFINITION 603063858F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212766 5',
mRNA sequence.
ACCESSION BI907284
VERSION BI907284.1 GI:16170094

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 754)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1534 row: d column: 15
High quality sequence stop: 750.
Location/Qualifiers
FEATURES
source
1..754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5212766"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/note="Vector: PCMV-SPOrt6; Site_1: NotI; Site_2: EcoRV
(destroyed); RNA source Leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research genetics tracking code 027. Note:
this is a NIH_MGC Library."
BASE COUNT 171 a 182 c 194 g 207 t
ORIGIN
alignment_scores:
Quality: 1236.00 Length: 251
Ratio: 4.944 Gaps: 1
Percent Similarity: 99.602 Percent Identity: 99.203
alignment_block:
US-09-776-865-2 x BI907284 ..
Align seg 1/1 to: BI907284 from: 1 to: 754
18 GLYGlyPheGlyLeuSerGlyArgArgSerLeuLeuGlyValAlaIle 34
|||||
6 GGGGGCTTGGGGCTGTGGGGCGGGCTCTCTTGTGCCAGGTGGGAG 55
|||||
34 rThrProAlaHisValAlaGlyValMetArgSerProValArgAspLeuAla 51
|||||
56 TACACCTGCTCAGCTAGGCGCTCATGAGGTCTCCGGTTCCGAGACTGGGCC 105
|||||
51 rGAsnAspGlyGluGlnSerThrAspArgThrProLeuLeuProGlyValA 67
|||||
106 GGAACGATGGGAGAGAGACGAGACCG.ACGCTCTTCTTACCGCGGCGCC 154
|||||
68 ProArgIaGluAlaAlaProValCysCysSerAlaArgTYrAsnLeuAl 84
|||||
155 CCAGGGCGCGAGCGCTCCAGTGTGCTGCTGCTGCTTACCAACTTAGC 204
|||||
84 AlIleAlaAlaPhePheGlyPhePheIleValTYrAlaLeuArgValAsn 101
|||||
205 AATTGTGGCTTTTGTGTTCTCATTTGTGTGATGCATTACGTGTGATC 254
|||||
101 euSerValAlaLeuValAlaSPMeValAspSerAsnThrThrLeuGluAsp 117
|||||

```

255 TCAGTGTTCGCTAGTGGATATGTTAGATTCAAAATACAACTTAGAAGAT 304
118 AsnArgThrSerLysAlaCysProGluHisSerAlaProIleLeuValHis 134
|||||
305 AATGAACTTCCAGGCGGTGCTCCAGACATTCTCTCCATTAAGTTCA 354
134 SHASnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnGly 151
|||||
355 TCATAAATCAAAACGGGTAAAGAGTACCAATGGCATGCAAGAACTCAAGAT 404
151 rPileLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGly 167
|||||
405 GGATTCCTCGTCTCTTTTATGCTACATCATCACACAGATTCTCGGA 454
168 GLTYrValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyIle 184
|||||
455 GGATATCTTCGACGAAATAGGGGGGAAATGCTGCTAGCATTTGGGAT 504
184 eLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeu 201
|||||
505 CCTGGCAGCTGCTGCTCCACCCCTGTCACCTCCATGCTGCAGATTAG 554
201 LysAlaGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGly 217
|||||
555 GAGTTGGACCACTCATTTGTAAGTACAGACACTAGAGAGAGAGAGG 604
218 ValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProProle 234
|||||
605 GTTACATTTCCAGCCATGCATGCCATGTGGCTTCTTGGGCTCCCTCT 654
234 uGluArgSerLysLeuLeuSerLysTyrAlaGlyAlaGlnLeuGly 251
|||||
655 TGAAGAAGAGAACTCTTACATTCATATGACAGACAGAGCTTGGA 704
251 hrValIle.SerLeuProLeuSerGlyIleIleGlyTyrTyrMetAsnTr 267
|||||
705 CAGTAATTTCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 753
267 p 267
754 G 754
seq_name: gb_est2:BI661062
seq_documentation_block: 801 bp mRNA linear EST 12-SRP-2001
LOCUS BI661062
DEFINITION 603304362F1 NIH_CGAP_Mam4 Mus musculus cDNA IMAGE:5350046 5',
mRNA sequence.
ACCESSION BI661062
VERSION BI661062.1 GI:15575298
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 801)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1889 row: 1 column: 15
High quality sequence stop: 799.
FEATURES
source
Location/Qualifiers
1..801

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5350046"
/clone_1lb="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furch, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
Library."
BASE COUNT 173 a 200 c 220 g 208 t
ORIGIN
alignment_scores:
Quality: 1115.00 Length: 257
Ratio: 4.705 Gaps: 4
Percent Similarity: 92.218 Percent Identity: 83.268
alignment_block:
US-09-776-865-2 x BI661062 ..
Align seg 1/1 to: BI661062 from: 1 to: 801
40 GlyValMetArgSerProValArgAspLeuAlaArgAsnAspGlyGlu 56
|||||
22 GGCCTTATGAGGCCGCCCTTCGCGGGTCCCGCGGAAACGCGATGAGGA 71
56 UserThrAspArgThrProLeuLeuProGlyAlaProAlaGluAla 73
|||||
72 GAGCTCGGACAGACCCCGCTCTGCGCGGCCCGCGGACGAGCAAGAGG 121
73 LaproValCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhe 89
|||||
122 CTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 171
172 GlyPhePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeu 106
|||||
106 LaspMetValAspSerAsnThrThrLeuGluAspAsnArgThrSerLys 123
222 GGACATGTGATTCATCAATCAATCAATCAATCAATCAATCAATCAAT 271
123 LacYsProGluHisSerAlaProIleLysValHisHisAsnGlnThrGly 139
272 AGTGTGGGAAACATTCGCCGCCCTTAAAGTCCACCATTCACACAGT 321
140 LysLysTyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerP 156
|||||
322 AAAAAGTACAAGTGGATGAGAACTCAAGGCGGATTCGCGCTCTTT 371
156 ePheTyrGlyTyrIleThrGlnIleProGlyLysTyrValAlaSerL 173
372 TTTTACGGCTACATCGTCAACCATTCGCGGGGTACATTCGCGCA 421
173 ysiIleGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
422 GGTGCGAGGGAAGCTGCTGCTGCGCTGGGCACTTACACACACAGT 471
190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeu 206
472 TTCACCTGTTCACACCGCTGCGCGAGACTTACGCTGGGATCTCGT 521
206 eValLeuArgAlaLeuGluGlyLeuGlyGlyValThrPheProAla 223
522 TGTCCTTGAAGCGCTGGAAGAGCTGGGAGAGGGTTTACGCTTCCAGCT 571
223 eHisAlaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeu 239

```

|||||
572 TGCAGCAGATGCTCTCCCTGGGCTCCCTCTGGAAGAAGCAGCTT 621
240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
|||||
622 CTTACCAATTCCTATGCGGAGACAGCTTGAGACATGATCTCACTTCC 671
256 OLeuSerGlyIleIleGlyTyrMetAsnTyrThrTyrValPheTyr 272
|||||
672 TCTTCCGGAATATATGCTATATGACTGAGCTTACGCTCTTATTC 721
273 PhePheGlyThrIleGlyIlePheThrPheLeuLeuTriPleTyrLeuV 289
|||||
722 TTT...GGTATAGTTGGAATGTGTGG...TTATTTATGATGTGTGATAG 765
289 a1SerAspThrProGln 294
|||||
766 TCAGTGATACACACAGA 782
seq_name: gb_est2:BE867611

```

```

seq_documentation_block: 676 bp mRNA linear EST 20-OCT-2000
LOCUS BE867611 601443127F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847279.5',
DEFINITION mRNA sequence.
ACCESSION BE867611
VERSION BE867611.1 GI:10316387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 676)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgraps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1AM9561 row: e column: 08
High quality sequence stop: 676.
Location/Qualifiers
source
1. 676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3847279"
/tissue_type="adrenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

```

```

FEATURES
source
1. 676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3847279"
/tissue_type="adrenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 145 a 171 c 184 g 176 t
ORIGIN

```

```

alignment_scores:
Quality: 1062.00 Length: 225
Ratio: 4.872 Gaps: 3
Percent Similarity: 96.889 Percent Identity: 95.111

```

```

alignment_block:
US-09-776-865-2 x BE867611 ..
Align seg 1/1 to: BE867611 from: 1 to: 676

```

```

10 ArgProValGlnProAla..ArgProGlyGlyPheGlyLeuSerGlyArg 25
|||||
5 CGACCCACGCGCTCCGCGCTCGCGCGCGCTTCGGGCTTCGGGCGG 54
26 ArgSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGlyValMe 42
|||||
55 CGCTCCCTCTCTCGCAGAGTGGCGAGTACACCTGCTCAGTAGGGGTCAT 104
42 tArgSerProValAlaArgAspLeuAlaArgAsnAspGlyGluLeuSerThr 59
105 GAGGCTCCGGTTGGAGACCTGGCCGGAGACGATGGCAGAGAGACAGC 154
59 sPargThrProLeuLeuProGlyAlaProAlaGlnAlaAlaProVal 75
|||||
155 ACCGACGCGCTCTTCTACCGGCGCGCCCGCGGAGGCGCGCTCCAGT 204
76 CysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyPheP 92
|||||
205 TGCTGCTCTGCTCGTTACACTTACATTTGCGCTTTTGGTTCTT 254
92 e1leValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAspMetV 109
|||||
255 CATGTGTATGCAATTCATGCTGAATCTGAGTGTGCTTACGTAGATATAG 304
109 a1AspSerAsnThrThrLeuGlnAspAsnArgThrSerValAlaCysPro 125
305 TAGATTCAATACAACTTATAGAGTATATAGACTTCCAAAGCATGTCCA 354
126 GluHisSerAlaProIleLeuValHisHisAsnGlnThrGlyLysLysT 142
|||||
355 GAGCATTCCTGCCCATTAAGATTCAATCAATCAAAAGGCTAGAGAGTA 404
142 rGlnThrPaspAlaGlnThrGlnGlyTyrIleLeuGlySerPhePheTyr 158
405 CCAATGGATGCAGAACTCAAGATGATGATTCCTGCTTCTTTGGTAT 454
159 GlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerLysile.. 174
|||||
455 GGCATCATCATCACACAGATTCCTGAGAGATATGTGGCACAACAAATACG 504
175 GlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeuTh 191
|||||
505 GGGGGGAAATGCTGCTGATGGATTCCTTGGCATGCTGCTCCTCAC 554
191 rLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleVal 208
555 CCGTCTTCTCCATTCGCTGCAGATTAGAGATTGACCATCTATTGTAC 604
208 euArgAlaLeuGlnGlyLeuGlyGlnGlyValThrPheProAlaMetHis 224
|||||
605 TCAGAGCAGCTAGAGAGACTAGAGAGGAGGTGTACATTTCCAGCATGAT 654
225 AlaMetTrpSerSerTrpAla 231
|||||
655 GCCATGTGCTCTCTTGGGCT 675
seq_name: gb_est2:BI860521
seq_documentation_block: 736 bp mRNA linear EST 10-OCT-2001
LOCUS BI860521 603386787F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772.5',
DEFINITION mRNA sequence.
ACCESSION BI860521
VERSION BI860521.1 GI:16001268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 736)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M12008 row: m column: 21
High quality sequence stop: 713.
Location/Qualifiers
1. 736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:5395772"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 216 a 142 c 147 g 231 t
ORIGIN

alignment_scores:
Quality: 1024.50 Length: 224
Ratio: 5.022 Gaps: 2
Percent Similarity: 91.071 Percent Identity: 88.839

alignment_block:
US-09-776-865-2 x B1860521 ..

Align seg 1/1 to: B1860521 from: 1 to: 736

```

120 ThisSerlysaIacysProgluHisSerAlaProIleuValHisHis 136
1 ACTTCCAGGCGTCCAGACATTCGCTCCCAATAAAGTTCATCTAA 50
136 ngInThGlyLysLysTyrGlnTrpAspAlaGluThrGlnGlyTrpIle 153
51 TCAACGCGTAAGAAATACCAATGGATGCAAACTCAAGATGATTC 100
153 euGlySerPhePheTyrGlyTyrIleLeuThrGlnIleProGlyTyr 169
101 TCGGTCCTTTTATGCTACATCATCACAGATTCCTGGAGATAT 150
170 ValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyIleLeu 186
151 GTTCCGACGAAANAGGGGAAATCTCTAGATTTGGGATCCTTG 200
186 yThAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuGlyVal 203
201 CACGTGCTCTCACCTGTTCACTCCATGCTGACGATTTAGAGATTG 250
203 LyrProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGlyValThr 219
251 GACCACCATTTGACTCAGACACTAGAACGACTAGAGAGGGGTACA 300
220 PheProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGlu 236
301 TTTCACACCATGCATGCCATGTGCTTCTTGGGCTCCCTCTTGAAG 350
236 gSerLysLeuLeuSerLysSerTyrAlaGlyAlaGlnLeuGlyThrVal 253
351 AAGCAAACTTCTAGCATTCATATGACAGACACAGCTTGGACAGTAA 400
253 leSerLeuProLeuSerGlyIleIleCysTyrTyrMetLeuThrPheTyr 269
401 TTTCTCTTCCCTTCTGGAATAATTTGCTACTATATGAAATGGACTTAT 450

```

```

270 ValPheTyrPhePheGlyThrIleGlyIlePheThrPheLeuLeuTrpIle 286
451 GTCTTCTACTT.TTGGTACTATTGGAATATTTGGTTCTTTCTGTGAT 499
286 eTrpLeuValSerAspThrProGlnLysHisLysArgIleSerHisTyrG 303
500 CTGGTTAGTTAGTGACACACCACAAACACAGAGATTTCCCATATG 549
303 LuYsgIuTyrIleLeuSerSerLeuArgGlnGlnLeuSerSerGlnLys 319
550 AAAAGGATTCATTTTCATTCATTAAGAAATCAG..... 584
320 SerValProTrpValProIleLeuLysSerLeuProLeuTrpAlaIle 336
585 .....GTATGGAATTGG.. 596
336 lValAlaHisPheSerTyrAsn 343
597 .....ACACATTTTAAATAAT 614
seq_name: gb_est2:BF676817
seq_documentation_block:
LOCUS BF676817 626 bp mRNA linear EST 21-DEC-2000
DEFINITION 602084380F1 NIH_MGC_83 Homo sapiens CDNA clone IMAGE:4248811 5',
mRNA sequence.
ACCESSION BF676817
VERSION BF676817.1 GI:11950712
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 626)
NIH-MGC http://mgi.nci.nih.gov/
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1067 row: o column: 20
High quality sequence stop: 598.
Location/Qualifiers
1. 626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:4248811"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATCTAGAGCGCGCGCGCGCATG-3' and 3' adaptor
sequence: 5'-ATCTAGAGCGCGCGCGCGCGCATG-3' (30)BN-3'.
Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 157 a 126 c 123 g 220 t
ORIGIN


alignment_scores:  
Quality: 991.00 Length: 189  
Ratio: 5.243 Gaps: 0


```

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-776-865-2 x BF676817 ..

Align seg 1/1 to: BF676817 from: 1 to: 626

```

348 ThrLeuLeuThrLeuLeuProThrTyMetLysGluIleLeuArgPheas 364
|||||
2 ACTTATTGACATTATTCCTACTATATGAAAGAGATCCCTAGGTTCAA 51
|||||
364 nValIngluAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTyrPL 381
|||||
52 TGTTCACAGAGATGGCTTTTATCTTCATTCCTATTAGGCTCTGGT 101
|||||
381 euCyMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTyrP 397
|||||
102 TATGTATGATCCTGTCTGCTCAAGCTGCTGACAAATTTAAAGGCAAAATGG 151
|||||
398 AsnPheSerThrLeuCyValArgArgIlePheSerLeuIleGlyMetI 414
|||||
152 AATTTTCAACTTATGTGTGGAGAAATTTTACCTTATAGAGATGAT 201
|||||
414 eGlyProAlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyrS 431
|||||
202 TGGACCTCAGATATCCTGTAGCTGTGCTTCATGTGCTGTGATTAAT 251
|||||
431 erLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCys 447
|||||
252 CTTTGCCGCTGTCTTCACTATATCAACAACACTGGAGGCTTTTGGC 301
|||||
448 SerSerGlyPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaG 464
|||||
302 TCTTCTGATTTAGCATCAACATCTGGATATTCCTCTGTATGATCGG 351
|||||
464 YIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetValG 481
|||||
352 TATCTCTCTGGGCAACAAATACATTTGCCATATTCAGAAATGGTTG 401
|||||
481 lYProValIleAlaIleSerLeuThrProAspAsnThrValGlyLysIuPr 497
|||||
402 GGCCCGCATTTGCTAAAGTCTGACCCCTGATTAACACATGTGGAGAAATGG 451
|||||
498 GlnThrValPheTyrIleAlaAlaIleAsnValPheGlyAlaIlePhe 514
|||||
452 CAACCGGTTCTATATGCTGCTCTATTAATGTTTGGGCCATTTT 501
|||||
514 ePheThrLeuPheAlaLysGlyIleValGlnAsnTyrPalaLeuAsnAspH 531
|||||
502 CTTTACACTATTCGCCCAAGGTGAGTACAAAACTGGGCTCTCATATGATC 551
|||||
531 lSHISGLYHISARGHIS 536
|||||
552 ACCATGGACACGACAC 568

```

seq_name: gb_est2:BI697765

seq_documentation_block:

LOCUS BI697765 711 bp mRNA linear EST 18-SEP-2001

DEFINITION 603346838F1 NCI_CGAP_Mam2 Mus musculus cDNA IMAGE:5374407 5', mRNA sequence.

ACCESSION BI697765

VERSION BI697765

KEYWORDS BI697765.1 GI:15660394

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 711)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

plate: LLAM1953 row: c column: 16

High quality sequence stop: 707.

FEATURES

source

location/Qualifiers

1..711

/organism="Mus musculus"

/strain="FVB/N-3"

/db_xref="taxon:10090"

/clone="IMAGE:5374407"

/clone_lib="NCI_CGAP_Mam2"

/tissue_type="tumor, biopsy sample"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 147 a 191 c 199 g 174 t

ORIGIN

alignment_scores:

Quality: 991.00 Length: 236

Ratio: 4.546 Gaps: 2

Percent Similarity: 92.373 Percent Identity: 83.051

alignment_block:

US-09-776-865-2 x BI697765 ..

Align seg 1/1 to: BI697765 from: 1 to: 711

```

33 AlaserThrProAlaHis.Val.....GlyValMetArgSerProValA 47
|||||
6 GCACGACGCGGCGACATCTCTCCAGCGGTGTGAGCCCTCGCTTC 55
|||||
47 TgAspLeuAlaArgAsnAspGlyGluSerThrAspArgThrProLeu 63
|||||
56 GGGGCGCGCGGGAACGACGATGAGGAGACTCGGAGACGCCCGCTC 105
|||||
64 LeuProGlyAlaProArgAlaGluAlaAlaProValCysSerSeraIaR 80
|||||
106 CTGCGCGGCGCGCGGACAGCCGAGCGGCTCCAGTGTGCTCTGCTGCG 155
|||||
80 GTTAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValIyrAlA 97
|||||
156 GTACAACTTAGCGATTTTGGGCTTGTGCTGTCTCTCTCTATGCT 205
|||||
97 euArgValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThr 113
|||||
206 TACGGGTGAACCTGAGTCTTCTGCTAGCGACATGATGATTTCAATACA 255
|||||
114 ThrLeuGluAspAsnArgThrSerLysAlaCysProGluHisSeraIaPr 130
|||||
256 ACTCTGACTGATTAATAGAACGCTTAAGAGAGTGTCCGGAACATTCGCC 305
|||||
130 oIlleValHisHisAsnGlnThrGlyLysTyrGlnTrrpAspAlag 147
|||||
306 CATAAAGTTCACCAATCACACAGGTAAAGATACAGGTGGAGTCAG 355
|||||
147 lufThrGlnGlyTrrPileLeuGlySerPhePheTyrGlyTyrIleIthr 163
|||||
356 AAACCTCAAGGSGTGTCTGGCTC.TTTTTHACGGCTACATGTCTACC 404
|||||
164 GlnIleProGlyGlyTyrValAlaSerLysIleGlyLysMetLeuLe 180
|||||
405 CAGATTCGCCGSGGTGATGATGTCAGCAGGAGGTCGAGGAAGATGCTGCT 454

```

alignment block:
US-09-776-865-2 x BF539146 ..

Align seg 1/1 to: BF539146 from: 1 to: 895

```
25 ArgArgSerLeuLeuCysGlnValAlaSerThrProAlaHisValGlyAla 41
||||| :|||||: |||||
4 CGAGCGCGGCACATCTGCTCT.....CCAGCGCT 32

41 lMetArgSerProValArgAspLeuAlaArgAsnAspGlyGluGluSer 58
||||| :||| ||| |||||
33 TAGGAGGCCCTGCTTCGGGGGTCGGCGGGAACGACGATGAGAGAGCT 82

58 lhrAspArgThrProLeuLeuProGlyAlaProArgAlaGluAlaPro 74
:|||| :|||||: |||||
83 CGGACGACGACCCCGCTCTGCGCGCGCCCGGACGACCGAGCGGCTCCA 132

75 ValCysCysSerAlaArgTyraSLeuAlaAlaLeuAlaPheGly 91
||||| :|||||: |||||
133 GGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181

91 ePheIleValTyraAlaLeuArgValAsnLeuSerValAlaLeuValAsp 108
||||| :|||||: |||||
182 CTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 230

108 eValAspSerAsnThrThrLeuGluAspAsnArgThrSerLysAlaCys 124
||||| :|||||: |||||
231 TGGTAAATTCAAATACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280

125 ProGlnHisSerAlaProIleLysValHisAsnGlnThrGlyLys 141
||||| :|||||: |||||
281 GCGGACATCTTCCGCCATAAAGTTCCACCAACATCACACAGTAAAA 330

141 sTyrglnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePhe. 157
||||| :|||||: |||||
331 GTACAGATGGGATGACAGAAACATCAAGGGTGATTCGCGCTTTTGTG 380

158 TyrglnTyrrIleIleThrGlnIleProGlyGlyTyrrValAlaSerLys 174
||||| :|||||: |||||
381 TACGGCTAATGCTGCACCCAGATTCGCCGGGTGATATTCACGACGGGT 430

174 eGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeu 191
||||| :|||||: |||||
431 CGGAGGGAAGCTGCTGCTGCGGCTCGGCGCATCTTAGCACCCTCGTTC 480

191 hrLeuPheThrProIleAlaAspLeuGlyValGlyProLeuIleVal 207
||||| :|||||: |||||
481 CCTGTTCACACCGCTGCGCGCAGACTTAGGCGCTGTGACTCTCGT. 529

208 LeuArgAlaLeuGluGlyLeuGlyGlyValThrPheProAlaMet 224
||||| :|||||: |||||
530 CTTAGGCGCCTGGAAGACTGGGAGAGGGTGTACCTTCCACCTTAGCA 579

224 sAlaMetTrpSerSerTrp. AlaProLeuGluArgSerLysLeuLeu 240
||||| :|||||: |||||
580 CGGCACTGTGCTCTCTGCGGCTCCCGCTCGAAGAAAGCAAGCTTCT 629

241 SerIleSerTyraGlyAlaGlnLeuGlyThrValIleSerLeuProle 257
||||| :|||||: |||||
630 ACCATTTCTATGCGGGAGCACAGCTTGAGACAGTATCATCTTCTCT 679

257 userGlyIleIleCysTyrrTyrrMetSerThr 267
||||| :|||||: |||||
680 TGCAGGAATTAATATGCTACTATATGAACGTG 710
```

seq_name: gb_est2:BF124137

seq_documentation block:

LOCUS	BF124137	904 bp	mRNA	linear	EST 24-OCT-2000
DEFINITION	601760838F1 NCL_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4024064 5',				
RNA sequence.	BF124137				

VERSION BF124137.1 GI:10963177
 EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 904)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabds-remail.nih.gov
 Tissue Procurement: Lotmar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LHAM9283 row: k column: 09
 High quality sequence stop: 680.

FEATURES
 source location/Qualifiers
 1..904
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:4024064"
 /clone_lib="NCI-CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Saliv; Site:2: Noli; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lotmar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 193 a 214 c 276 g 221 t
 ORIGIN

alignment_scores:
 Quality: 909.00 Length: 233
 Ratio: 4.170 Gaps: 0
 Percent Similarity: 93.562 Percent Identity: 84.549

alignment_block:
 us-09-776-865-2 x BF124137 ..

Align seg 1/1 to: BF124137 from: 1 to: 904

40 GYValMetArgSerProValArgAspLeuAlaArgAsnAspGlyGluG1 56
 |||||
 24 GCGGTATGAGGCCCTCGTCCGGGTCGCGGGAACGACGATGAGGA 73
 |||||
 56 userThrasparGthrProLeuLeuProGlyAlaProArgAlaGluAlaA 73
 |||||
 74 GAGCTGGAGACAGACCCGCTCTCCGCGCGGACGACGACGACGCG 123
 |||||
 73 laProValCysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePhe 89
 |||||
 124 CTCGAGTGTGCTGCTCTCGGTACCACTTACCATTTGCGGCTTCTGT 172
 |||||
 90 GlyPhePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuVa 106
 |||||
 173 GGTTCCTCTGTTCTATGCTTACGGGTGAACCTGAGTGT.GCGTTAGT 221
 |||||
 106 lAspMetValaspSerAsnThrThrLeuGluAspAsnArgThrSerLysA 123
 |||||
 222 GGACATGTAGATTCATCAATACACTGACGTATATAGAACGCTTAGAG 271
 |||||
 123 lacysProGluHisSerAlaProIleLysValHisHisGlnThrGly 139
 |||||
 272 AGTGTGCGGAACATTCTGCCCCCATTAAGTTACACATACACACAGGT 321
 |||||

140 LysLysTyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPh 156
 |||||
 322 AAAAGTACAGATGGATGACAGAAACTCAAGGTGATTCCTGCTCTT 371
 |||||
 156 ePheTyrGlyTyrIleIleIlePheGlnIleProGlyGlyTyrValAlaSerL 173
 |||||
 372 TTTTACCGGTACATCGTACCCAGATTCGCCGTGGTACATTTGCCACA 421
 |||||
 173 ySileGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
 |||||
 422 GGGTGGAGGGAAGCTGCTGGGCTGGGCACTTATGGAACCTCCCTC 471
 |||||
 190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuI 206
 |||||
 472 TTCACCCGTTCACACCGCTGGCGCACACTTAGCGGTGACCTCTGT 521
 |||||
 206 eValLeuArgAlaLeuGluGlyLeuGlyGlyValThrPheProAlaM 223
 |||||
 522 TGTGCTTAGACCGCTGGAAGACTGGAGAGGTGTACGTTCACAGCTA 571
 |||||
 572 TGCAAGC.ATGTGCTCTCTCGGCTCCCTCGGAAGAAGCAAGCTT 620
 |||||
 240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
 |||||
 621 CTACACATTTCTCTATGCGGAGACACAGTT.GGACACAGATCTACCTCC 669
 |||||
 256 oLeuSerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyr 272
 |||||
 670 TCTTTC.GGAATTAATATGTACTATATGAATGACCTACCTCTCTCAT 717
 |||||

seq.name: gb_hlc:BC023331

seq_documentation_block:
 LOCUS BC023331 1531 bp mRNA linear HTC 05-FEB-2002
 DEFINITION Mus musculus, clone IMAGE:4500938, mRNA.
 ACCESSION BC023331
 VERSION BC023331.1 GI:18497348
 KEYWORDS HTC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1531)
 AUTHORS Strausberg, R.
 DIRECT SUBMISSION
 SUBMITTED (05-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabds-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: angdbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

REMARK
 COMMENT
 JOURNAL

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 29 row: f column: 13
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: no poly-a-tail

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="IMAGE:4500938"
/tissue_type="Eye, retina, mouse strain C57BL/6
/clone_lib="NIN.MGC_94"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
BASE COUNT      371 a      458 c      434 g      268 t
ORIGIN

```

alignment_block:
US-09-776-865-2 x BC023331/rev .

53 AspGylgluInuSerThrAspArgThr.....ProLeuLeuPro.. 65
 |||||::: || ::||| |||:::|
 1360 GACGGCGGCCAGTGACCACGACACTCGGACCCGCTGTGTGGACTG 1311

66GlyAlaProArgAlaGluAlaAlaProValCysCysSera 79
 ||| |||||
 1310 CACCTGCTTTGGCCTCCCTCGT 1289

79 laa^gly^rasn^leu^ala^lle^uala^phe^phe^gly^phe^phe^lle^val^tyr 95
 ||||| ::|||::: ::||| || :::
 1288 ..CGTACATCATCGCATCATGAGCGTCTGGGTTTCTGTATCAGCTTT 1241

```

96 AlALEuargValAsnLeuSerValAlaleuValaspmetValaspSerAs 112
   :::::| | | | | | | | | | | | | | | | | | | | | | | | | |
1240 GGCATCCGCTGCACACCTGGGGCGTGGCCATCTGTCCATGGTCAACAACAG 1191

```

```

112 nThrLeuGluAspAsnArgThrSerLysAlaCysProGluHisSerA 129
      :|||||      :::::::::::
1190 CACAACC.....CACCGTGGGG 1174

```

```

129 laProleIysValHisHisangInThrGlySlySTyrgInTrpasp 145
:: ::: ||||::: ||| :::::||||||
1173 GCCACGTGGTGGTGCAGAAAGCCACAG.....TTCAACTGGGAT 1136

```

146 ALAGLThrcInGlyTrpIleLeucIGlySerPhePheTyrGLYTrlIleI 162
 ||||| ||| ||| |||||||||||:::|||||:::
 1135 CCAGAGACTGTGGCCCTCATACATGGCTCTTTTTCGTGGGGCTACATTGT 1088

```

162 eThurcInIleProGlyGlyTyrValAlaSerLysIleGlyGlyLysMetL 179
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1085 CACTCAGATTCTCGAGAGATTATCTGCCAAAAATTGCAGCCAAACAGGG 1036

```

```

179 euleuGlypheGlyIleleuGlyThrAlaValleuThrleuPheThrPro 195
      :::::::::::::::::::::::::::::::::::::::::::: |||
1035 TCCTTGcCTTGCCATGCTGGcCTACCTCCACcCTTAACATGTTGATCCCT 986

```

196 IleAlaIAspLeuGlyValGlyProLeuIleValIleuArgAlaLeuGI 212
 ||||| :: ||| ::||| ::||| |||:
 985 TCAGCAGCCCGCGTTCATATGGCTGTGCATCTTCGGAGGATCTTCA 936

212 uGIylLeuGIyGlucLyValThrpheProAlaMethHisAlaMetTrpSers 229
::||||| | ||||| ::::| ||||| | ||::::| ||||| :
935 GGGATTGGTGGAGGGGCTACATACCTCGTCGTGCCTATGGCAATCCGAAGA 886

229 eRTpAlaProProlGluArgSerLysLeuSerIleSerTyrAla 245
::| | | | | | | | | | | | | | | | | | :: : : : : : : : :
885 AATGGCGCCCTTCCTTCAACGCCACTGCCTCCGCAACCGAACAGCCTTTTGTG 926

246 GlyAlaGlnLeuGlyThrValIleSerLeuProLeuSerGlyIleIleCys 262

835 GGTTCCTATGCTGGGGCGGTGGTTGCCATGCCCTTGGCTGGGGTCCTTGT 786

785 GCAGTATTCAGGATGGAGTCTGTCTTCTATGTCTATGCGCAGCTTCGGGA 736

735 TCCTTTGGTACCTGTTCTGGTTCCTTGTCTCCTATGAGTCAACCGGCACTG 686

685 CACCCAGCATCTCTGAGGAGGAGCGCAATACATTGAGGATGCCATCGG 636

635 GGAGAGCGCCAAAGCTCATGAACCCCTGTACGAAGTTTAAACACACACCCCTGGA 586

585 GCGGCTCTTACGTCATGCCCGTCTATGCCATCATCGTTGGCACTTT 536

[illegible]

357 LYSGLILELEUVALGPNESNVALGINGLUALSNGLYPNELEUSERL 3/4
 ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
 485 TGAAGAGTGTTCGGCTTTGAGATCAGCAAGTGGGGCTGCTTCGGCCC 436

386
390
394
435

TGCTCACCTTGTCATGACCATCATCGTACCATTGGAGGCCAGATCGCT
|||||::||| :::: ||||| |
eurotyrleucylsertrpleucysmethleuserylinalala
386

391 aspasnleuargalalystipasnpheserThrleucysValArgArgI 407
 385 GACCTTTTGGCGAGTCGTACATAATGTCCACTACCAACGTCGGAAGCT 336

407 ephaserleuileglymettileglyProAlaValPheUValAlaA1ag 424
::: : |::: :|::: :|::: :|::: :|
335 CATGAAC TCGGGGCTTTCGGGATGGAAGCCACGCTGCTGCTGCTGC 286

424 LysphnelelglycysasptyrserleualaValAlaphleuthrileser 440
||::: ||: :::::||:::||:::||:::||:::
285 GATAC..TCGCACTCCAAGGGCGTGCCATCTCCTTCTG6GTGGCT 239

441 ThrThrLeuglyglyphcycsserSerglypheSerIleasnHisLeuas 457
:: ::::::| |::: |||||:::| |
238 GTGGGCTTCAGTGCGTTTGCATCTCTGGGTAAACGTGAACCACTTGGGA 189

457 pilealaProserTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPhea 474
188 CATGCCCTCGCTATGCCAGCATCTTGATGGGCATTTCCAAATGGCGTGG 139

474 laThrIleProGlyMetValGlyProValIleAlaLysSerLeuThrPro 490
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
138 GCACACTGTCTGGGATGGTGTGCCCATCATTCGTGGTGCAATGACCAAG 89

491 AspasnThrValGlyGluTrpGlnThrValPheTyrIleAlaAlaIle 507
:::| | | | | | | | | | | | | | | |
88 CACAAGACGGCGGAGGAGTGCCAGTAGCTGTTCCTCATAGCCTCCCTGGT 39

```

507 easnvalphecglyalailephecThrLeupheala 519
      :::: |||:::|||||::: ::|||
38 GCACTACGGCGGTCTCATCTCTATGGGGCTTTGCT 2

```

```
name: gb_est2:BG541099
documentation_block:
```

LOCUS	BC554109	893 bp	mRNA	linear	EST 03-APR-2001
DEFINITION	602570265F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4694680 5', mRNA sequence.				

ACCESSION BG541099
 VERSION BG541099.1 GI:13533332
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 893)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: LCM1519 row: a column: 17
 High quality sequence stop: 549.
 Location/Qualifiers
 1..893
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4694680"
 /clone_lib="NIH-MGC-77"
 /lab_host="DH10B (rT1 phage-resistant)"
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggccatctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCGCGCCGACGATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."
 BASE COUNT 236 a 169 c 214 g 274 t
 ORIGIN
 alignment_scores:
 Quality: 900.00 Length: 325
 Ratio: 4.072 Gaps: 9
 Percent Similarity: 68.000 Percent Identity: 63.077
 alignment_block:
 US-09-776-865-2 x BG541099 ..
 Align seg 1/1 to: BG541099 from: 1 to: 893
 203 GlyProLeuIleValIleuArgAlaLeuGluGlyGluGlyValTh 219
 1 GGACCACTCATTTGACTCAGAGCACTAGAGAGGCTGTAC 50
 219 rPheProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGln 236
 51 ATTTCACGCCATGCATGCATGTGTCTTCTTGCGCTCCCTTTGAAA 100
 236 rGSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrVal 252
 101 GAACCAACACTTTCATGATTCATATGAGAGACAGACTGGACAGTA 150
 253 IleSerLeuProLeuSerGlyIleIleCysTyrTrpMetAsnTrpThrTy 269
 151 ATTTCCTCTCTCTCTCTGGAATATTTGCTACTATATGAAATTCAGCTTA 200
 269 rValPheTrpPhePheGlyThrIleGlyIlePheTrpPheLeuLeuTrp 285
 201 TGTCTTACTT.TTGGTACTATTGGAATATTGTTGTTCTTGG...TTGG 246

286 IleTrpLeuValSerAspThrProGlnLysHisLysArgIleSerHisTy 302
 247 ATCTGGTTAGTTAGTACACACACAAAAACAAAGAAATTCCTCCATTA 296
 302 rGlnLysGluTyrIleLeuSerSerLeuArgAsnGlnLeuSerSerGlnL 319
 297 TGAAGAAGGAATACATCTTTCATCATTAAGAAATCAG..... 333
 319 ysserValProTrpValProIleLeuLysSerLeuProLeuTrpAlaIle 335
 333 333
 336 ValValAlaHisPheSerTyrAsnTrpThrPheTyrThrLeuThrIle 352
 333 333
 352 uLeuProThrTyrMetLysGluIleLeuArgPheAsnValGlnGlnAsn 369
 334AATG 337
 369 LysPheLeuSerSerLeuProTyrLeuGlySerTrpLeuCysMetIleLeu 385
 338 GGTTTTATCTTCATTCCTTATTTAGCCTTGGTTATGATATCATCTG 387
 386 SerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSerThrL 402
 402 eucysValArgArgIlePheSerLeuIleGlyMetIleGlyProAlaVal 418
 438 TATGTGTGCGAGAAATTTAGCCTTATAGCAATGATTTGGACCTGCAGTA 487
 419 PheLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaValAl 435
 488 TTCTGTGAGTCTGCTGGCTTCATTTGGGTGATTAATCTTTGGCCGTCT 537
 435 alphaLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGlyPhe 452
 538 TTCTAATATATCAACAACACT...GGAGCCTTTGGTCTCTTCGGAATTA 584
 452 erIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeuLeuGln 468
 585 GATCCACCCATTTGGGATATG..... 605
 468 yIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValIleA 485
 606GTGCTTCCACACCTTTGGGAGAAATGGGA 636
 485 IalysSerLeuThrProAspAsnThrValGlyLutrginThrValPhe 501
 637 CCGGGTCTCTT.....ATGGTGGGGGGAATTAAGTCT..... 668
 502 TyrIleAlaAlaAlaIleAsnValPheGlyAlaIlePheThrLeuPhe 518
 669GGTGGCCACTTCTTATCACAATTT 691
 518 eAlaLysGlyLysValGln 524
 692 TGGCAAAAGCTGAGGCTCCAA 710
 seq_name: gb_est1:BB613552
 seq_documentation_block:
 LOCUS BB613552 682 bp mRNA linear EST 26-OCT-2001
 DEFINITION BB613552 RIKEN full-length enriched, 10 day neonate skin Mus
 ACCESSION BB613552
 VERSION BB613552.1 GI:16454120
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

LOCUS BB537525 645 bp mRNA linear EST 26-OCT-2001
DEFINITION BB537525 RIKEN full-length enriched, 0 day neonate eyeball Mus musculus CDNA clone E130006F23 3' similar to AJ387747 Homo sapiens
ACCESSION BB537525
VERSION BB537525.2 GI:16446752
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 645)
AUTHORS Arakawa,T., Carlinici,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Himemoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,T., Okido,T., Saito,C., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
TITLE Unpublished (2001)
JOURNAL
COMMENT On Jul 31, 2000 this sequence version replaced gi:9593025.

```
seq_name: gb_est1:AL597124
seq_documentation_block:
LOCUS AL597124 499 bp mRNA linear EST 14-AUG-2001
DEFINITION DKFZP313H1911.t1 313 (synonym: hlcc2) Homo sapiens cDNA clone
ACCESSION AL597124
VERSION AL597124.1 GI:15154820
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 499)
AUTHORS Koehler,K., Beyer,A., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Koehler,K., Beyer,A., Mewes,W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Koehler K
MIPS
Am Klopferstritz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by BMFZ (Biomedical Research Center at the Charité,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No sl sequence available.
This clone (DKFZP313H1911) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
FEATURES
source
1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP313H1911"
/clone_1ib="313 (synonym: hlcc2)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTR10plex2; Site_1: SfiIA; Site_2: SfiIB;
CDNA-collection"
BASE COUNT 110 a 129 c 133 g 126 t 1 others
ORIGIN
alignment_scores:
Quality: 862.00 Length: 165
Ratio: 5.256 Gaps: 0
Percent Similarity: 99.394 Percent Identity: 98.182
Alignment_block:
US-09-776-865-2 x AL597124 ..
Align seg 1/1 to: AL597124 from: 1 to: 499
9 ProAArgProValGlnProAlaArgProGlyGlyPheGlyLeuSerGlyArg 25
|||||
3 CCCC GCCCGG GTCAGACAGCTCGGCGGCGTTCGGGCGTGTGGGCGG 52
|||||
25 GAGSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGlyValM 42
|||||
53 GCGCTCCCTCTCTCTGCGAGGTGGGAGTACACCTGCTCAGGTAGCGCTCA 102
|||||
42 eArgSerProValArgAspLeuAlaArgAsnAspGlyGlnGluSerThr 58
|||||
103 TGAAGTCTCGGTTTCGAGACTGCGCCGGAACGATGGCGAGAGAGACAGC 152
|||||
59 AspArgThrProLeuLeuProGlyAlaProArgAlaGlnAlaAlaProVa 75
|||||
153 GACCGCAGCGCTTCTTACCGGGCGCCCAAGGGCGGAGCGCTCCAGT 202
|||||
75 lGyCySerAlaArgTyrAsnLeuAlaIleLeuAlaPheGlyPheP 92
|||||
```

```
203 GTGCTGCTGCTGCTGCTTACCACTTAGAATTTTGGCCCTTTTGGTTCCT 252
|||||
92 helLeValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAspMet 108
|||||
253 TCATTTGTTGATGCTTACGTTTATCTGACTGTTCGGCTTAGTGGATATG 302
|||||
109 ValAspSerAsnThrThrLeuGluAspAsnArgThrSerTyrAlaCysPr 125
|||||
303 GTAGATTCAATATACACTTTAGAGATTAATAGAACTTCCAAAGCGGTGCC 352
|||||
125 OGluHisSerAlaProIleLeuValHisAsnGlnThrGlyLysIst 142
|||||
353 AGACCAATTCACCTCCCAATAAAGTTCAATCAATCAACACGGGTAAAGAT 402
|||||
142 YrgInTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTyr 158
|||||
403 ACCAATGGAGTGCAGAACTCAAGATGATTCGGTTCCTTTTAT 452
|||||
159 GLYTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerLys 173
|||||
453 GCGTACATNATCACACAGATTCCTGAGAGATGTCGCCAGCAANA 497
|||||
seq_name: gb_est1:AA833297
seq_documentation_block:
LOCUS AA833297 518 bp mRNA linear EST 23-FEB-1998
DEFINITION ud05d09.t1 Soares.NMPu Mus musculus cDNA clone IMAGE:1434257 5'
similar to WP:C36C10.2 CE00105 SODIUM/PHOSPHATE TRANSPORTER ;, mRNA
sequence.
ACCESSION AA833297
VERSION AA833297.1 GI:2907025
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 518)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LINTL; contact the
IMAGE Consortium (info@image.lnlnl.gov) for further information.
MGI:918325
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 444.
Location/Qualifiers
FEATURES
source
1..518
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1434257"
/clone_1ib="Soares_NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus; and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 123 a 105 c 121 g 169 t
ORIGIN

alignment_scores:
Quality: 802.00 Length: 172
Ratio: 4.981 Gaps: 0
Percent Similarity: 93.605 Percent Identity: 85.465

alignment_block:
US-09-776-865-2 x AA833297 ..

Align seg 1/1 to: AA833297 from: 1 to: 518

```
275 GIYThrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerAs 291
||| ::::||| |||||:|||||:|||||:|||||:|||||:
1 GGTATAGTGTGAATGTCTGCTTTATTTATGATGTGATAGTCAGTCA 50
291 pThrProGlnLysHisLysArgIleSerHisTrpGlnLysGlnTrpIleL 308
|||||:|:| ||||| |||||:|||||:|||||:|||||:
51 TACACCAGAACTCACAGACAACTCCCATATGAAAAAGAAATACATTG 100
308 euSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValProTrpVal 324
:::|||||:|:| |||||:|||||:|||||:|||||:
101 TTTCAATCAATAAAAAATCAGCTTCTTCGACAGAGTGCGCTGGGCG 150
325 ProIleLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSe 341
|||||:|:| |||||:|||||:|||||:|||||:
151 TCCATCTTGAACTCAGCTGCACCTTGGGCAATTGGTGACACATTCTC 200
341 rTyrAsnTrpIlePheTrpThrLeuLeuThrLeuLeuProThrTyrMetL 358
|||||:|:| |||||:|||||:|||||:|||||:
201 CTACAACTGGTCTTTTACACCTTATGTGACGCTACTGCCAACTTATATGA 250
358 ySGLuIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeu 374
|||||:|:| |||||:|||||:|||||:|||||:
251 AGGAATCCCTAAGGTTCATGTTCAGAGAAAGGGGTTTTTATCTGCATTG 300
375 ProTyrLeuGlySerTrpLeuLysMetIleLeuSerGlyGlnAlaAlaAs 391
|||||:|:| |||||:|||||:|||||:|||||:
301 CCTTATTTGGCTGTGGTATGATGATCCCTGTGTGCAAGCCGCTGA 350
391 pAsnLeuArgAlaLysTrpAsnPheSerThrLeuLysValArgArgIleP 408
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
351 CTATTTAAGGCTCAAGTGAAGTCTTCAACTATTAAGCGTTGCGAGCATTT 400
408 heSerLeuIleGlyMetIleGlyProAlaValPheLeuValAlaIleGly 424
|||||:|:| |||||:|||||:|||||:|||||:
401 TTAGCCTCGTAGAAGGTGGCCCTGGGCTTCTTAGTCGCGGCTGGA 450
425 PheIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerTh 441
|||||:|:| |||||:|||||:|||||:|||||:
451 TTTATAGGCTGTGACTATTCTTGSCCGTTGCGTTCCTAACCATATCCAC 500
441 rThrLeuGlnGlyPhe 446
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
501 GACGCTGGAGGAGGCTTC 516
```

Tue Jul 16 08:11:13 2002

OM of: US-09-776-865-2 to: GenEmbl:* out-format : pfs
Date: Jul 15, 2002 3:35 PM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPRO.spool/US09776865/runat_15072002.0922208_27415/app_query.fasta_1.1159
-DB=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPX4=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTWRT=pfs
-NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-DEV=US09776865_@CGN1_1.5372 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:
Query: US-09-776-865-2
Query length: 536
Database: GenEmbl:*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 3807.270000

score_list:

Sequence	Strid	Orig	ZScore	Len	Documentation
gb.pat:AX207624	+	2836.00	4141.55	2,5e-222	2930 1 AX207624 Sequence 1 from Patent
gb.pr:AF244577	+	2836.00	4141.55	2,5e-222	2930 1 AF244577 Homo sapiens membrane
gb.pat:AX138494	+	2804.00	4096.26	8,4e-220	2512 1 AX138494 Sequence 2 from Patent
gb.pr:HS387747	+	2804.00	4096.26	8,4e-220	2512 1 AJ387747 Homo sapiens mRNA for
gb.com:AF240591	+	2770.50	4043.92	6,9e-217	3329 1 BC020961 Homo sapiens, soluble
gb.com:AF244578	+	2338.00	3410.44	1,3e-181	2844 1 AF244578 Ovis aries membrane g
gb.pat:AX207626	+	2338.00	3410.44	1,3e-181	2844 1 AX207626 Sequence 3 from Patent
gb.in:AY060776	+	1922.00	1438.35	9,4e-72	1841 1 AY060776 Drosophila melanogast
gb.hlg:AC014246	+	984.50	1395.08	2,4e-69	33444 1 AC014246 Drosophila melanogast
gb.in:AC023685	+	984.50	1377.13	2,4e-68	167928 1 AC023685 Drosophila melanogast
gb.in:AC023711	+	984.50	1369.84	6,1e-68	323461 1 AC023711 Drosophila melanogast
gb.in:AE003491	+	973.00	1401.86	1,0e-69	3982 1 AF271235 Rattus norvegicus dif
gb.ro:AF271235	+	971.00	1403.98	7,7e-70	2528 1 AF24864 Mus musculus vesicula
gb.pr:AF24864	+	968.00	1394.62	2,6e-69	3346 1 AB032435 Homo sapiens DNPI mRN
gb.pr:AB032435	+	942.00	1362.86	1,5e-67	2366 1 AB032436 Homo sapiens BNPI mRN
gb.pat:IA0028	+	942.00	1360.59	2,0e-67	2716 1 I40028 Sequence 1 from Patent
gb.pat:IA0029	+	942.00	1360.59	2,0e-67	2716 1 I40029 Sequence 3 from Patent
gb.pat:IA0220	+	942.00	1360.59	2,0e-67	2716 1 I40220 Sequence 1 from Patent
gb.pat:IA0221	+	942.00	1360.59	2,0e-67	2716 1 I40221 Sequence 3 from Patent
gb.pat:IA0222	+	942.00	1360.59	2,0e-67	2716 1 I40222 Sequence 1 from Patent
gb.pat:IA0223	+	942.00	1360.59	2,0e-67	2716 1 I40223 Sequence 3 from Patent
gb.pat:IA0224	+	942.00	1360.59	2,0e-67	2716 1 I40224 Sequence 1 from Patent
gb.pat:IA0225	+	942.00	1360.59	2,0e-67	2716 1 I40225 Sequence 3 from Patent
gb.pat:IA0226	+	942.00	1360.59	2,0e-67	2716 1 I40226 Sequence 1 from Patent
gb.pat:IA0227	+	942.00	1360.59	2,0e-67	2716 1 I40227 Sequence 3 from Patent
gb.pat:IA0228	+	942.00	1360.59	2,0e-67	2716 1 I40228 Sequence 1 from Patent
gb.pat:IA0229	+	942.00	1360.59	2,0e-67	2716 1 I40229 Sequence 3 from Patent
gb.pat:IA0230	+	942.00	1360.59	2,0e-67	2716 1 I40230 Sequence 1 from Patent
gb.pat:IA0231	+	942.00	1360.59	2,0e-67	2716 1 I40231 Sequence 3 from Patent
gb.pat:IA0232	+	942.00	1360.59	2,0e-67	2716 1 I40232 Sequence 1 from Patent
gb.pat:IA0233	+	942.00	1360.59	2,0e-67	2716 1 I40233 Sequence 3 from Patent
gb.pat:IA0234	+	942.00	1360.59	2,0e-67	2716 1 I40234 Sequence 1 from Patent
gb.pat:IA0235	+	942.00	1360.59	2,0e-67	2716 1 I40235 Sequence 3 from Patent
gb.pat:IA0236	+	942.00	1360.59	2,0e-67	2716 1 I40236 Sequence 1 from Patent
gb.pat:IA0237	+	942.00	1360.59	2,0e-67	2716 1 I40237 Sequence 3 from Patent
gb.pat:IA0238	+	942.00	1360.59	2,0e-67	2716 1 I40238 Sequence 1 from Patent
gb.pat:IA0239	+	942.00	1360.59	2,0e-67	2716 1 I40239 Sequence 3 from Patent
gb.pat:IA0240	+	942.00	1360.59	2,0e-67	2716 1 I40240 Sequence 1 from Patent
gb.pat:IA0241	+	942.00	1360.59	2,0e-67	2716 1 I40241 Sequence 3 from Patent
gb.pat:IA0242	+	942.00	1360.59	2,0e-67	2716 1 I40242 Sequence 1 from Patent
gb.pat:IA0243	+	942.00	1360.59	2,0e-67	2716 1 I40243 Sequence 3 from Patent
gb.pat:IA0244	+	942.00	1360.59	2,0e-67	2716 1 I40244 Sequence 1 from Patent
gb.pat:IA0245	+	942.00	1360.59	2,0e-67	2716 1 I40245 Sequence 3 from Patent
gb.pat:IA0246	+	942.00	1360.59	2,0e-67	2716 1 I40246 Sequence 1 from Patent
gb.pat:IA0247	+	942.00	1360.59	2,0e-67	2716 1 I40247 Sequence 3 from Patent
gb.pat:IA0248	+	942.00	1360.59	2,0e-67	2716 1 I40248 Sequence 1 from Patent
gb.pat:IA0249	+	942.00	1360.59	2,0e-67	2716 1 I40249 Sequence 3 from Patent
gb.pat:IA0250	+	942.00	1360.59	2,0e-67	2716 1 I40250 Sequence 1 from Patent
gb.pat:IA0251	+	942.00	1360.59	2,0e-67	2716 1 I40251 Sequence 3 from Patent
gb.pat:IA0252	+	942.00	1360.59	2,0e-67	2716 1 I40252 Sequence 1 from Patent
gb.pat:IA0253	+	942.00	1360.59	2,0e-67	2716 1 I40253 Sequence 3 from Patent
gb.pat:IA0254	+	942.00	1360.59	2,0e-67	2716 1 I40254 Sequence 1 from Patent
gb.pat:IA0255	+	942.00	1360.59	2,0e-67	2716 1 I40255 Sequence 3 from Patent
gb.pat:IA0256	+	942.00	1360.59	2,0e-67	2716 1 I40256 Sequence 1 from Patent
gb.pat:IA0257	+	942.00	1360.59	2,0e-67	2716 1 I40257 Sequence 3 from Patent
gb.pat:IA0258	+	942.00	1360.59	2,0e-67	2716 1 I40258 Sequence 1 from Patent
gb.pat:IA0259	+	942.00	1360.59	2,0e-67	2716 1 I40259 Sequence 3 from Patent
gb.pat:IA0260	+	942.00	1360.59	2,0e-67	2716 1 I40260 Sequence 1 from Patent
gb.pat:IA0261	+	942.00	1360.59	2,0e-67	2716 1 I40261 Sequence 3 from Patent
gb.pat:IA0262	+	942.00	1360.59	2,0e-67	2716 1 I40262 Sequence 1 from Patent
gb.pat:IA0263	+	942.00	1360.59	2,0e-67	2716 1 I40263 Sequence 3 from Patent
gb.pat:IA0264	+	942.00	1360.59	2,0e-67	2716 1 I40264 Sequence 1 from Patent
gb.pat:IA0265	+	942.00	1360.59	2,0e-67	2716 1 I40265 Sequence 3 from Patent
gb.pat:IA0266	+	942.00	1360.59	2,0e-67	2716 1 I40266 Sequence 1 from Patent
gb.pat:IA0267	+	942.00	1360.59	2,0e-67	2716 1 I40267 Sequence 3 from Patent
gb.pat:IA0268	+	942.00	1360.59	2,0e-67	2716 1 I40268 Sequence 1 from Patent
gb.pat:IA0269	+	942.00	1360.59	2,0e-67	2716 1 I40269 Sequence 3 from Patent
gb.pat:IA0270	+	942.00	1360.59	2,0e-67	2716 1 I40270 Sequence 1 from Patent
gb.pat:IA0271	+	942.00	1360.59	2,0e-67	2716 1 I40271 Sequence 3 from Patent
gb.pat:IA0272	+	942.00	1360.59	2,0e-67	2716 1 I40272 Sequence 1 from Patent
gb.pat:IA0273	+	942.00	1360.59	2,0e-67	2716 1 I40273 Sequence 3 from Patent
gb.pat:IA0274	+	942.00	1360.59	2,0e-67	2716 1 I40274 Sequence 1 from Patent
gb.pat:IA0275	+	942.00	1360.59	2,0e-67	2716 1 I40275 Sequence 3 from Patent
gb.pat:IA0276	+	942.00	1360.59	2,0e-67	2716 1 I40276 Sequence 1 from Patent
gb.pat:IA0277	+	942.00	1360.59	2,0e-67	2716 1 I40277 Sequence 3 from Patent
gb.pat:IA0278	+	942.00	1360.59	2,0e-67	2716 1 I40278 Sequence 1 from Patent
gb.pat:IA0279	+	942.00	1360.59	2,0e-67	2716 1 I40279 Sequence 3 from Patent
gb.pat:IA0280	+	942.00	1360.59	2,0e-67	2716 1 I40280 Sequence 1 from Patent
gb.pat:IA0281	+	942.00	1360.59	2,0e-67	2716 1 I40281 Sequence 3 from Patent
gb.pat:IA0282	+	942.00	1360.59	2,0e-67	2716 1 I40282 Sequence 1 from Patent
gb.pat:IA0283	+	942.00	1360.59	2,0e-67	2716 1 I40283 Sequence 3 from Patent
gb.pat:IA0284	+	942.00	1360.59	2,0e-67	2716 1 I40284 Sequence 1 from Patent
gb.pat:IA0285	+	942.00	1360.59	2,0e-67	2716 1 I40285 Sequence 3 from Patent
gb.pat:IA0286	+	942.00	1360.59	2,0e-67	2716 1 I40286 Sequence 1 from Patent
gb.pat:IA0287	+	942.00	1360.59	2,0e-67	2716 1 I40287 Sequence 3 from Patent
gb.pat:IA0288	+	942.00	1360.59	2,0e-67	2716 1 I40288 Sequence 1 from Patent
gb.pat:IA0289	+	942.00	1360.59	2,0e-67	2716 1 I40289 Sequence 3 from Patent
gb.pat:IA0290	+	942.00	1360.59	2,0e-67	2716 1 I40290 Sequence 1 from Patent
gb.pat:IA0291	+	942.00	1360.59	2,0e-67	2716 1 I40291 Sequence 3 from Patent
gb.pat:IA0292	+	942.00	1360.59	2,0e-67	2716 1 I40292 Sequence 1 from Patent
gb.pat:IA0293	+	942.00	1360.59	2,0e-67	2716 1 I40293 Sequence 3 from Patent
gb.pat:IA0294	+	942.00	1360.59	2,0e-67	2716 1 I40294 Sequence 1 from Patent
gb.pat:IA0295	+	942.00	1360.59	2,0e-67	2716 1 I40295 Sequence 3 from Patent
gb.pat:IA0296	+	942.00	1360.59	2,0e-67	2716 1 I40296 Sequence 1 from Patent
gb.pat:IA0297	+	942.00	1360.59	2,0e-67	2716 1 I40297 Sequence 3 from Patent
gb.pat:IA0298	+	942.00	1360.59	2,0e-67	2716 1 I40298 Sequence 1 from Patent
gb.pat:IA0299	+	942.00	1360.59	2,0e-67	2716 1 I40299 Sequence 3 from Patent
gb.pat:IA0300	+	942.00	1360.59	2,0e-67	2716 1 I40300 Sequence 1 from Patent
gb.pat:IA0301	+	942.00	1360.59	2,0e-67	2716 1 I40301 Sequence 3 from Patent
gb.pat:IA0302	+	942.00	1360.59	2,0e-67	2716 1 I40302 Sequence 1 from Patent
gb.pat:IA0303	+	942.00	1360.59	2,0e-67	2716 1 I40303 Sequence 3 from Patent
gb.pat:IA0304	+	942.00	1360.59	2,0e-67	2716 1 I40304 Sequence 1 from Patent
gb.pat:IA0305	+	942.00	1360.59	2,0e-67	2716 1 I40305 Sequence 3 from Patent
gb.pat:IA0306	+	942.00	1360.59	2,0e-67	2716 1 I40306 Sequence 1 from Patent
gb.pat:IA0307	+	942.00	1360.59	2,0e-67	2716 1 I40307 Sequence 3 from Patent
gb.pat:IA0308	+	942.00	1360.59	2,0e-67	2716 1 I40308 Sequence 1 from Patent
gb.pat:IA0309	+	942.00	1360.59	2,0e-67	2716 1 I40309 Sequence 3 from Patent
gb.pat:IA0310	+	942.00	1360.59	2,0e-67	2716 1 I40310 Sequence 1 from Patent
gb.pat:IA0311	+	942.00	1360.59	2,0e-67	2716 1 I40311 Sequence 3 from Patent
gb.pat:IA0312	+	942.00	1360.59	2,0e-67	2716 1 I40312 Sequence 1 from Patent
gb.pat:IA0313	+	942.00	1360.59	2,0e-67	2716 1 I40313 Sequence 3 from Patent
gb.pat:IA0314	+	942.00	1360.59	2,0e-67	2716 1 I40314 Sequence 1 from Patent
gb.pat:IA0315	+	942.00	1360.59	2,0e-67	2716 1 I40315 Sequence 3 from Patent
gb.pat:IA0316	+	942.00	1360.59	2,0e-67	2716 1 I40316 Sequence 1 from Patent
gb.pat:IA0317	+	942.00	1360.59	2,0e-67	2716 1 I40317 Sequence 3 from Patent
gb.pat:IA0318	+	942.00	1360.59	2,0e-67	2716 1 I40318 Sequence 1 from Patent
gb.pat:IA0319	+	942.00	1360.59	2,0e-67	2716 1 I40319 Sequence 3 from Patent
gb.pat:IA0320	+	942.00	1360.59	2,0e-67	2716 1 I40320 Sequence 1 from Patent
gb.pat:IA0321	+	942.00	1360.59	2,0e-67	2716 1 I40321 Sequence 3 from Patent
gb.pat:IA0322	+	942.00	1360.59	2,0e-67	2716 1 I40322 Sequence 1 from Patent
gb.pat:IA0323	+	942.00	1360.59	2,0e-67	2716 1 I40323 Sequence 3 from Patent
gb.pat:IA0324	+	942.00	1360.59	2,0e-67	2716 1 I40324 Sequence 1 from Patent
gb.pat:IA0325	+	942.00	1360.59	2,0e-67	2716 1 I40325 Sequence 3 from Patent
gb.pat:IA0326	+	942.00	1360.59	2,0e-67	2716 1 I40326 Sequence 1 from Patent
gb.pat:IA0327	+	942.00	1360.59	2,0e-67	2716 1 I40327 Sequence 3 from Patent
gb.pat:IA0328	+	942.00	1360.59	2,0e-67	2716 1 I40328 Sequence 1 from Patent
gb.pat:IA0329	+	942.00	1360.59	2,0e-67	2716 1 I40329 Sequence 3 from Patent
gb.pat:IA0330	+	942.00	1360.59	2,0e-67	2716 1 I40330 Sequence 1 from Patent
gb.pat:IA0331	+	942.00	1360.59	2,0e-67	2716 1 I40331 Sequence 3 from Patent
gb.pat:IA0332	+	942.00	1360.59	2,0e-67	2716 1 I40332 Sequence 1 from Patent
gb.pat:IA0333	+	942.00	1360.59	2,0e-67	2716 1 I40333 Sequence 3 from Patent
gb.pat:IA0334	+	942.00	1360.59	2,0e-67	2716 1 I40334 Sequence 1 from Patent
gb.pat:IA0335	+	942.00	1360.59	2,0e-67	2716 1 I40335 Sequence 3 from Patent
gb.pat:IA0336	+	942.00	1360.59	2,0e-67	2716 1 I40336 Sequence 1 from Patent
gb.pat:IA0337	+	942.00	1360.59	2,0e-67	2716 1 I40337 Sequence 3 from Patent
gb.pat:IA0338	+	942.00	1360.59	2,0e-67	2716 1 I40338 Sequence 1 from Patent
gb.pat:IA0339	+	942.00	1360.59	2,0e-67	2716 1 I40339 Sequence 3 from Patent
gb.pat:IA0340	+	942.00	1360.59	2,0e-67	2716 1 I40340 Sequence 1 from Patent
gb.pat:IA0341	+	942.00	1360.59	2,0e-67	2716 1 I40341 Sequence 3 from Patent
gb.pat:IA0342	+	942.00	1360.59	2,0e-67	2716 1 I40342 Sequence 1 from Patent
gb.pat:IA0343	+	942.00	1360.59	2,0e-67	2716 1 I40343 Sequence 3 from Patent
gb.pat:IA0344	+	942.00	1360.59	2,0e-67	2716 1 I40344 Sequence 1 from Patent
gb.pat:IA0345	+	942.00	1360.59	2,0e-67	2716 1 I40345 Sequence 3 from Patent
gb.pat:IA0346	+	942.00	1360.59	2,0e-67	2716

67 aProRgaLagLualAaIaProValCyCySerAlaArgTyraSnLeuA 84
 443 CCCAGGGGCGAAGCGGCTCCAGTGGCTGCTGCTGCTTACCACTTAG 512
 84 lAlLeuAlaPhePheGlyPhePheIleValTyraAlaLeuArgValasn 100
 513 CAATTGTCCTTTTGGTTCTTCATTGCTATGATGCAATTAGCTGTGAAT 562
 101 LeuSerValAlaLeuValAspMetValAspSerAsnThrThrLeuGluAs 117
 563 CTGAGTGTGGCTTGGCTTAGTGATGTGGTACATTCAAAATCAACTTTAGAGA 612
 117 PAsnArgThrSerLysAlaCySProGluHisSerAlaProIleLysValH 134
 613 TAAATGAACCTTCCAGCGGCTGTCAGAGACATTCTGCCATTAAGAAAGTTC 662
 134 lSHIsAsnGlnThrGlyLysLysTyrglnTrpAspAlaGluThrGlnGly 150
 663 ATCAATAACTCAACGGGTAAAGATCAATGGATGCAAGAACTCAAGAGA 712
 151 TrpIleLeuGlySerPhePheTyrglyTyrlleIleThrGlnIleProgl 167
 713 TGGATTTCGGTTCCTTTTATATGCTACATCATCATCACAGATTCCTGG 762
 167 yGlyTyraValAlaSerLysIleGlyGlyLysMetLeuLeuGlyPheGlyI 184
 763 AGGATATGTGTCAGCAAAATAGGGGGAAAAATGCTGCTAGATTTGGGA 812
 184 lLeuGlyThrAlaValleuThrLeuPheThrProIleAlaAlaLysLeu 200
 813 TCTTGGGACGTGCTGTCTCACCCCTGTCACCTCCATTGCTGCAGATTYA 862
 201 GlyValGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGlyGluG 217
 863 GGAGTTGGACCACTCATTTGACTGACAGACACTAGAAGACTAGAGAGAGG 912
 217 yValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProL 234
 913 TGTTCATTTCCAGCCATGATGATGCTGTGCTTGGGCTCCCTCCCTC 962
 234 euGluArgSerLysLeuSerLysIleSerTyraGlyAlaGlnLeuGly 250
 963 TTGAAGAAGCAAACTTCTAGCATTTGTAAGCAGAGACAGCTTGGG 1012
 251 ThrValIleSerLeuProLeuSerGlyIleIleCySerTyrMetAsnTr 267
 1013 ACAGTAATTTCTCTCTCTCTCTGTAATATTGCTACTATATGAAATG 1062
 267 pThrTyraValPheTyPhePheGlyThrIleGlyIlePheTrpPheLeu 284
 1063 GACTTATGTCTTCTACTTTTGGTACTATGGAATATTTGGTTCTTT 1112
 284 euTrpIleTrpLeuValSerAspThrProGlnIleHisLysArgLysLe 300
 1113 TGTGATGTGGTGTAGTAGTGACACACACAAACAAAGAAATTTCC 1162
 301 HisTyrglyLysGlyTyrlleLeuSerSerLeuArgAsnGlnLeuSer 317
 1163 CATATGAAAAGAAATACATTTCTTCAATCAATTAAGAAATCAAGCTTCTTC 1212
 317 rGlnLysSerValProTrpValProIleLeuLysSerLeuProLeuTrpA 334
 1213 ACAGAAAGTCACTGGCGGTGGTACCATTTTAAATCCCTGCCACTTTGGG 1262
 334 lAlleValValAlaHisPheSerTyraAsnTrpThrPheTyThrLeuLeu 350
 1263 CTATCGTAGTGGACACATTTCTTCAACACTGACATTTTAACTTATG 1312
 351 ThrLeuLeuProThrTyMetLysGluIleLeuArgPheAsnValGlnG 367
 1313 ACATATATGCTCTACTTATATAGAGAGATCCTTAAGTTCAATGTTCAAGA 1362
 367 uAsnGlyPheLeuSerSerLeuProTyrlleuGlySerTrpLeuGlyMet 384

1363 GAATGGTTTTTATCTTCATTCAGCTTATTTAGGCTTGGTTAGTATGA 1412
 384 lLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
 1413 TCTGTCGTGGTCAAGCTGCTGACAAATTTTAAGGGCAAAATTTGCA 1462
 401 ThrLeuCyValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
 1463 ACTTTATGTGTTGGCAGAAATTTTATAGCTTATAGAAATGATGGACCTGC 1512
 417 aValPheLeuValAlaAlaGlyPheIleGlyCyAspTrpSerLeuAlav 434
 1513 AGTATTCCTGCTAGTGTGCTGCTTCATGCTGATGATATCTTTGGCCG 1562
 434 aAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCySerSerGly 450
 1563 TTGCTTTCCTAATATATCAACAACTGGAGGCTTTTGGCTTCTGGA 1612
 451 PheSerIleAsnHisLeuAspIleAlaProSerTyraLysIleLeuLe 467
 1613 TTTAGCATCAACCATCTGATATTCCTTCCTGATGCTGATATCTCTCT 1662
 467 uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValI 484
 1663 GGGCATCAAAATACATTTGCCACTATTCAGGAATGGTGGCCCGTCA 1712
 484 lAlaLysSerLeuThrProAspAsnThrValGlyGluTrpGlnThrVal 500
 1713 TTGCTTAAAGTGTGACCCCTGATTAACACTGTTTGGAGAAATGGCAACCGTG 1762
 501 PheTyrlleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrLe 517
 1763 TTTCTATATGCTGCTGCTATTAATGTTTGGTGCATTTCTTTTACACT 1812
 517 uPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisGlyH 534
 1813 ATTGGCCAAAGTGAAGTACAAAACTGGGCTCAATGATCATCATGATGAC 1862
 534 lArgHis 536
 1863 ACAGACAC 1870
 seq_name: gb-pr:AF244577
 seq_documentation_block: 2930 bp mRNA linear PRI 06-AUG-2000
 LOCUS AF244577 Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
 DEFINITION Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
 ACCESSION AF244577
 VERSION AF244577.1 GI:9719373
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 (bases 1 to 2930)
 Fu,C., Barhan,S., Ceteanu,N.D., Lloyd,S.R., Yan,H.-P.,
 Carter,C.E., Shi,E., Venkov,C., Yakes,M.F., Page,D.L. and
 Hellerqvist,C.G.
 Identification of a novel membrane protein from mammalian cells
 that interacts with the anti-pathoangioenic compound CM101
 Unpublished
 2 (bases 1 to 2930)
 Fu,C., Barhan,S., Ceteanu,N.D., Lloyd,S.R. and Hellerqvist,C.G.
 Direct Submission
 Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School
 of Medicine, 23rdPierce, Nashville, TN 37232-0146, USA
 FEATURES
 source 1..2930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2930
 /gene="HP59"
 gene

CDS
263..1873
/gene="HP59"
/note="similar to Na⁺ inorganic phosphate transporter"
/codon_start=1
/product="membrane glycoprotein HP59"
/protein_id="AA097769.1"
/db_xref="GI:9719374"
/translation="MAAGAMPAPRPVOPAPRGFGSLGSRSLTCQVASTPAWGMRS
PVRLARDGEESTDRPILPGLPAGRAEAPVCCSARYNALIAEFGFPIYALRVNLS
VALVADMSNTLTEDNRTSKRCPENHSAPIKVHNQTKKKYQMDAETGGMLISPFYGY
ITITIPGQYASKIGGKMLGFLGTAVALTLFTPIAADLGVGLIYRALGEGEY
TFPAMHAWSSWAPLESKLISISYGAQLGTVISLPSGITCYMMWVYEFET
IGIFWFLIMLVSDTPQKRIKISHYEKEYILSLSRNLSQKSVPVVPIILKSLPWA
IVVAHESYNTFTYTLTLPLPYMKEILRFNOENGFLSLPYLSMLCMILSGOALN
LRAKMFSTLCVRIFSLIGMIPAVFLVAGFVGCQYSLAVAFILITSLIGGFCSSG
FSTIHDIAPSYACILIGITPFTATIGMVPVIAKSLTPNTVGENQYVYIAALIN
VFGLIFLFLKGEVQNMALNDHGRH"

BASE COUNT 729 a 698 c 681 g 822 t
ORIGIN

alignment_scores:
Quality: 2836.00 Length: 536
Ratio: 5.291 Gaps: 0
Percent similarity: 100.000 Percent identity: 100.000

alignment_block:
US-09-776-865-2 x AF244577 ..

Align seg 1/1 to: AF244577 from: 1 to: 2930

1 MetAlaIaGlAlaMetThrProArgProValGlnProAlaArgpr 17
|||||
263 ATGCGCGCGGGGCGATGACACCGCCGCCGCTCCACGCGACTGGGCC 312
|||||
17 oglyglyPheglyLeuSerGlyValArgSerLeuLeuGlyValAla 34
|||||
313 CGGGGGCTCGGGCTCGGGCGCGGCTCCCTTCTGCGCAGTGCGCA 362
|||||
34 erThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 50
|||||
363 GTACACCTGCTCAGTAGGGCTCAGAGCTCCGGTTCGAGACCTGGCC 412
|||||
51 ArgAsnAspGlyGlnGluSerThrAspArgThrProLeuLeuProGlyAla 67
|||||
413 CGGAGCGATGGCGAGAGACGACGACGCGCTCTTCTACCGGGCGC 462
|||||
67 aProArgAlaGlnAlaAlaProValCysCysSerAlaArgTyAsnLeuAla 84
|||||
463 CCCAGCGGCGGAGCGCGCTCCAGTGTGCTGCTCGTTACAACTTAG 512
|||||
84 laLeuLeuAlaPhePheglyPhePheIleValIleValArgValAsn 100
|||||
513 CAATTTGGCCCTTTTGTGTTCTTCATGTGTATCATTAAGTGTGAAT 562
|||||
101 LeuSerValAlaLeuValAspMetValAspSerAsnThrThreugLnuAs 117
|||||
563 CTAGAGTGTGCGTTAAGTATGATGATGATTCAAATACAACTTAAAGA 612
|||||
117 pasnArgThrSerLysAlaCysProGlnHisSerAlaProIleLysValH 134
|||||
613 TAATAGAACTTCCAGGCGGTCCAGAGCATTCGCTCCCATAAAGTTC 662
|||||
134 lshHisnGlnThrGlyLysLysTyrgIntTrraspAlaGlnThrGlnGly 150
|||||
663 ATCATATCAACAGGGTAAAGTACCAATGGGATGCAGAACTCAAGGA 712
|||||
151 TrpIleLeuGlySerPhePheTyrgTyrlleIleThrGlnIleProGly 167
|||||
713 TGCATTCGCGTTCCTTTTATGCGTACATCATACACAGATTCTCTGG 762
|||||
167 yglyTyValAlaSerLysIleGlyLysMetLeuLeuGlyPheglyI 184
|||||

763 AGGATATGTTGCCAGCAAAATAGGGGGAAMATGCTGTAAGATTGGGA 812
|||||
184 leuLeuGlyThrAlaValleuThrLeuPheThrProIleAlaAlaAspLeu 200
|||||
813 TCCTTGGCACTGCTGCTCCACCCCTTCTCAGTCCCATTCGTCAGATTAA 862
|||||
201 GlyValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGln 217
|||||
863 GGAGTGGAGCAACCTATGCTAGTACAGACAGTAAAGAGACTAGAGAGAG 912
|||||
217 yValThrPheProAlaMetHisAlaMetTrpSerSerThrPalProPro 234
|||||
913 TGTTACATTTCCAGCCATCCATCCATGTGCTCTTCTGGGCTCCGCC 962
|||||
234 engLuarSerLysLeuLeuSerIleSerTyraGlnValGlnLeuGly 250
|||||
963 TTGAAGAGCAACACTTCTTACATTGATGACAGAGACAGCTTGGG 1012
|||||
251 ThrValIleSerLeuProLeuSerGlyIleIleCysTyrrTyrrMetAsn 267
|||||
1013 ACAGTAAATTTCTCTCTCTCTGGAATATTTGCTACTATATGAATG 1062
|||||
267 pThrTyrrValPheTyrrPhePheglyThrIleGlyIlePheThrPheLeu 284
|||||
1063 GACTTATGCTCTTCTACTTTTGGTACTATGGAATATTTGGTCTCTT 1112
|||||
284 eutRplIleThrLeuValSerAspThrProGlnHisLysArgIleSer 300
|||||
1113 TGTGATCTGTTAGTTAGTACACACCAACAAACACAGAAATTTCC 1162
|||||
301 HisTyrgLysGlnTyrrIleLeuSerSerLeuArgAsnGlnLeuSer 317
|||||
1163 CATTATGAAGAAGATATACATTTCTTCATCATTAAGAATACAGCTTCTTC 1212
|||||
317 rgLynsSerValProThrValProIleLeuLysSerLeuProLeuTrp 334
|||||
1213 ACAGAACTAGTCCCGTGGTACCCATTTAAATCCCTGCCACTTTGGG 1262
|||||
334 laLeuValAlaHisPheSerTyrrAsnTrpPheTyrrThrLeuLeu 350
|||||
1263 CTATCTAGTGTGACACTTTCTTACACTGACCTTTTATATATTTATG 1312
|||||
351 ThrLeuLeuProThrTyrrMetLysGlnIleLeuArgPheAsnValGln 367
|||||
1313 ACATTTATGCTCTTATATGAAAGAGATCCTTAAGTTCAATGTTCAAGA 1362
|||||
367 uAsnGlyPheLeuSerSerLeuProTyrrLeuGlySerTrpLeuCysMet 384
|||||
1363 GAATGGGTTTATTTCTTCATTTGCTTATTTAGGCTCTTGTTATGATGA 1412
|||||
384 leuLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
|||||
1413 TCCTGTGCTGCAAGGTGCTGACAAATTTAAGGGCAAAATGGAATTTTGA 1462
|||||
401 ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
|||||
1463 ACTTATGTGCTTGCGAAGATTTTACCTTATAGAAATGATGAGACTGCG 1512
|||||
417 aValPheLeuValAlaAlaGlyPheIleGlyCysAspTyrrSerLeuAla 434
|||||
1513 AGTATTCCTGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1562
|||||
434 alaAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
|||||
1563 TTGCTTCTCTAATATATCAACAACCTGAGAGGCTTTGCTCTTCTGGA 1612
|||||
451 PheSerIleAsnHisLeuAspIleAlaProSerTyrrAlaGlyIleLeuLe 467
|||||
1613 TTAGATATCAACCATCTGATGATGCTCTGCTGCTGCTGCTGCTGCTGCT 1662
|||||
467 uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValI 484
|||||
1663 GGGCATACAAATATACATTTGCCACTATTCAGGAATGTTGGGCCGCTCA 1712
|||||

```

484  LealaIysSerLeuThrProAspAsnThrValGlyIuTrpGlnThrVal 500
      |||||||
1713  TTGCTAAAAGCTGACCCCTGATACACTGTGGAGAAATGCAACCTG 1762
      |||||||
501  PheTyrIleAlaIleAlaIleAsnValPheGlyAlaIlePhePheThrLe 517
      |||||||
1763  TTCATATATGCTGCTGCTATTAATGTTTGTGTCATTTCTTTACACT 1812
      |||||||
517  uPheAlaIysGlyIuValGlnAsnTrpAlaLeuAsnAspHisHsGlyH 534
      |||||||
1813  ATTCGCCAAAGTGAAGTACAAACTGGGCTCTCAATGATCACCATGGAC 1862
      |||||||
534  ISARGHIS 536
      |||||||
1863  ACAGACAC 1870

seq_name: gb_pat:AX138494

seq_documentation_block:
LOCUS AX138494 2512 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 2 from Patent EP1069184.
ACCESSION AX138494
VERSION AX138494.1 GI:14274389
KEYWORDS
SOURCE
ORGANISM human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 2512)
REFERENCE
AUTHORS
TITLE Human anion transporter gene implicated in salt disease and
JOURNAL lysosomal static acid transport
PATENT: EP 1069184-A 2 17-JAN-2001;
AKZO NOBEL N.V. (NL)
FEATURES
source location/Qualifiers
            1..2512
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 612 a 585 c 566 g 749 t
ORIGIN
alignment_scores:
            Quality: 2804.00 Length: 535
            Ratio: 5.261 Gaps: 1
Percent Similarity: 99.626 Percent Identity: 99.252

alignment_block:
US-09-776-865-2 x AX138494
Align seg 1/1 to: AX138494 from: 1 to: 2512

```

```

85  LeuAlaIlePhePheGlyPhePheIleValTyrAlaLeuArgValAsnLeu 101
      |||||||
401  TTTTGGCTTTTGGTTTCTTCATTTGTATGATTCACGTGTAATCTG 450
      |||||||
102  SerValAlaLeuValAspMetValAspSerAsnThrThrLeuGlnAspAs 118
      |||||||
451  AGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
      |||||||
118  nArgThrSerIysAlaCysProGlnHisSerAlaProIleIysValHisH 135
      |||||||
501  TAGAAGCTTCCAGAGCGGTGTCAGAGCATTCGCTCCATTAAGTTATC 550
      |||||||
135  ISAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGlnTrpGlnGlyTyr 151
      |||||||
551  ATATCAACGGGTAGAGATGATGATGATGATGATGATGATGATGATGATG 600
      |||||||
152  IleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyG 168
      |||||||
601  ATTCGCGTTCCTTTTATGCTACATCATCACACAGATTCCTGGAGG 650
      |||||||
168  YTYrValAlaSerIysIleGlyGlyLysMetLeuLeuGlyPheGlyIle 185
      |||||||
651  ATATGTTGCCAGCAAAATAGGGGGGAAATGCTGCTAGGATTTGGATTC 700
      |||||||
185  euGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuGly 201
      |||||||
701  TTGGCACTGCTGCTCCTCACCCCTGTCACCTCCATGCTGCAGATTTAG 750
      |||||||
202  ValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGlyIle 218
      |||||||
751  GTTGACACCACTCATTTGACTCAGAGCACTAGAAAGACTAGACAGCGGT 800
      |||||||
218  lThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProLeuG 235
      |||||||
801  TACATTTCCAGCCATGCAATGCAATGCTCTGCGGCTCCCTCTTG 850
      |||||||
235  IuArgSerIysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThr 251
      |||||||
851  AAAAAGCAAACTTCTTAGCATTTTCATATGCAAGAGACAGAGCTTGGACA 900
      |||||||
252  ValIleSerLeuProLeuSerGlyIleIleCysTyrTyrMetAsnTrpH 268
      |||||||
901  GTAATTTCTCTTCTCTTCTGGAATATTTGCTACTATATGAAATGAGAC 950
      |||||||
268  rTyrValPheTyrPhePheGlyThrIleGlyIlePheTrpPheLeuLeu 285
      |||||||
951  TTATGCTCTCTACTTTTGTGACTATGGAATATTTGTTGTTCTTTGT 1000
      |||||||
285  rPileTrpLeuValSerAspThrProGlnIysHisIysArgIleSerHis 301
      |||||||
1001  GGATCTGCTTAGTATGATGACACACACACAAACAAACAGAGATTTCCAT 1050
      |||||||
302  TyrGlnLysGlnTyrIleLeuSerSerLeuArgAsnGlnLeuSerSerG 318
      |||||||
1051  TATGAAAGGAATCATTTCTTCATCATTAAGAAATCAGCTTCTCTTCACA 1100
      |||||||
318  nLysSerValProTrpValProIleLeuLysSerLeuProLeuTrpAla 335
      |||||||
1101  GAAATCAGTGGCGGTGCTACCATTTTAAATCCCTCCACTTGGGGCTA 1150
      |||||||
335  LeValAlaIleHisPheSerTyrAsnTrpThrPheTyrTrpLeuLeuThr 351
      |||||||
1151  TCGTAGTTCACACTTTTCTTACACTGGACTTTTATCTATTATTTAGACA 1200
      |||||||
352  LeuLeuProThrTyrMetLysGlnIleLeuArgPheAsnValGlnGlnAs 368
      |||||||
1201  TTATTTGCTCTACTTATATGAGAGATCTTAAGGTTCAATGTTCAAGAGAA 1250
      |||||||
368  nGlyPheLeuSerSerLeuProTyrIleGlySerTrpLeuCysMetIleL 385
      |||||||
1251  TGGGTTTTTATCTTCATTTGCTTATTTAGGCTCTTGGTTTATGATGATCC 1300
      |||||||
385  euSerGlyGlnAlaIleAspAsnLeuArgAlaLysTrpAsnPheSerThr 401

```

```
|||||
1301 TGTCTGGTCAAGCTGCTGCACATTTAAGGGCAAAATGGAATTTTCAACT 1350
402 LeuGyValArgArgIlePheSerLeuIleGlyMetIleGlyProAlaVala 418
1351 TTATGTGTTCCGAGAAATTTTTCCTTAAGGAATGATTGGACCTGCAGT 1400
418 IPhelValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaVala 435
1401 ATTCTGTGAGCTGCTGGCTTTCATTTGGCTGTATTTCTTTGGCCGTTG 1450
435 IAPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGlyPhe 451
1451 CTTTCTTAACATATATCAACAACACAGGAGCGCTTTGCTCTTGGAATTT 1500
452 SerLeuAsnIleLeuAspIleAlaProSerTyrAlaGlyIleLeuLeuG 468
1501 AGCATCAACCACTGTGATATGCTCTTCGTATGCTGTATCCTCTGCGG 1550
468 YIleThrAsnThrPheAlaThrIleProGlyMetValGlyProAlaIle 485
1551 CATCAACAATACATTTGCCACATATCCAGAAATGTTGGCCCGCATTTG 1600
485 IAlaYSerLeuThrProAspAsnThrValGlyGlyIleThrPheAlaPhe 501
1601 CTAAGAAGCTGCACCCCTGATACACTGTTCGAGAAATGGCAACCGTTC 1650
502 TyrIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrLeuP 518
1651 TATATGTCGCTGCTATTAATGTTTTTGGTCCATTTCTTTACACTATTT 1700
518 eAlaIaYSerGlyGlyValGlnAsnThrPAlaLeuAsnAspHisGlyHis 535
1701 GCCCAAGGTGAAGTAAACAACGAGGCTCTCAATGATCATCAGTGAACACA 1750
535 IGHIS 536
1751 GACAC 1755

seq_name: gb_HSA387747
LOCUS HSA387747 2512 bp mRNA linear PRI 10-DEC-1999
DEFINITION Homo sapiens mRNA for sialin.
ACCESSION AJ387747
VERSION AJ387747.1 GI:6562532
KEYWORDS sialin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2512)
AUTHORS Verheijen, F.W., Verbeek, E., Aul, N., Beeren, C.E., Havelaar, A.C.,
Jooose, M., Peltonen, L., Aul, P., Galjaard, H., van der Spek, P.J. and
Mancini, G.M.
TITLE A new gene, encoding an anion transporter, is mutated in sialic
acid storage diseases
JOURNAL Nat. Genet. 23 (4), 462-465 (1999)
MEDLINE 20047778
REFERENCE 2 (bases 1 to 2512)
AUTHORS Verheijen, F.W.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1999) F.W. Verheijen, Erasmus University, Dept.
Clinical Genetics, P.O. Box 1738, 3000 DR Rotterdam, NETHERLANDS
FEATURES
source
location/Qualifiers
1..2512
/organism="Homo sapiens"
/db_xref="taxon:9606"
271..1758
/codon_start=1
/product="sialin"
/protein_id="CA862540.1"
/db_xref="GI:6562533"
```

```
/translation="MRSPVRLARNDEESTDRTPLLGPAPRAEAAPVCCSARYNLAI
LAFGEFIYVALRVNLISVALVDWNSNTLEDNRTSACEBAPIKVHNQTKYO
WDAEFGFIYVALRVNLISVALVDWNSNTLEDNRTSACEBAPIKVHNQTKYO
VGPLYLIRALEGEGEYTPPAMAMMSNAPPLERKSLSISAGAOIGTIVISLPSG
ITCYMMWTVYTFEFGTIGLFWFLMWLWVSDTPQKHRIKSHYEKEYTISLNQLSS
OKSPWVPIKRSPLMAIIVAHPSYMWTFYTLPLPYKIELRNVQNGRLSSIP
YLSMWLCMISGQADNLRANKNFTLCVARIISLGMIPAVFLVAAGFICGDYSLA
VAFELTSTIGFGCSGFSINHLDIAPSVAGILIGTINFPATIPGHWGPVIAKSLP
NTVEMQTVYIAAINVREGATFEFLFACGEVNMALNDHGRH"
```

BASE COUNT 612 a 585 c 566 g 749 t

ORIGIN

alignment_scores:

Quality:	2804.00	Length:	535
Ratio:	5.261	Gaps:	1
Percent Similarity:	99.626	Percent Identity:	99.252

alignment_block:

us-09-776-865-2 x HSA387747 ..

Align seg 1/1 to: HSA387747 from: 1 to: 2512

```
4 glyAlaMetThr.....ProProArqProValGlnProAlaArgProG 18
|||||
151 GGGGGGGGTGTCATCGCCCGCCCGCCGGTCCAGCCAGCTCGGCCGG 200
18 yGlyPheGlyLeuSerGlyArgArgSerLeuLeuGlyGlnAlaIleSert 35
|||||
201 GGGCTTGGGGGTGGTGGGGCGGGCGCTCCCTTCCTGCGCAGGTTGGAGTA 250
35 hrProAlaHisValGlyAlaMetArgSerProValArgAspLeuAlaArg 51
|||||
251 CACCTGCTCAGCTAGCGCTCATGAGCTCTCCGTTCCGAGACCTGGCCGG 300
52 AsnAspGlyGlnGluSerThrAspArgThrProLeuLeuProGlyAlaPr 68
|||||
301 AACGATGGCGAGAGAGACGACGACCGCAGCCCTTCTTACGGGGCGCC 350
68 CARGAlaGluAlaAlaProValCysCysSerAlaArgTyrAsnLeuAlaI 85
|||||
351 ACGGGCCGACAGCCGCTCCAGTGTCTGCTGCTGCTGCTGCTGCTGCTG 400
85 IeLeuAlaPhePheGlyPhePheIleValTyrAlaLeuArgValaAsnLeu 101
|||||
401 TTTTGCGCTTTTGTGTTCTTCATTTGTATGCTATACGTGTGAATCTG 450
102 SerValAlaLeuValaIleAspMetValaIleAspSerAsnThrThrLeuGluAspAs 118
|||||
451 AGTGTGGCTGTAGTGAATGATGATCAATCAACACTTATGAGAGATTA 500
118 nArgThrSerIysAlaCysProGlnHisSerAlaProIleIysValaHis 135
|||||
501 TAGAAGCTTCCAGAGCGGTCCAGACCATTCGCTCCATAAAGTTCTTC 550
135 ISAsnGlnThrGlyIysIysTyrGlnTyrPASPAlaGlnThrGlnGlyTyr 151
|||||
551 ATATCAACAACGGGTAAAGATACCAATGGATGCGAAGAACCTCAAGATGG 600
152 ILeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyG 168
|||||
601 ATTCTGCGTCTCTTTTATGGCTACATCAACAGATTCGCGAGG 650
168 YTyrValAlaSerIysIleGlyGlyMetLeuLeuGlyPheGlyIleL 185
|||||
651 ATATGTTGCCAGCAAAATATGAGGGAATGCTGTAGATTGGGATTC 700
185 euGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuGly 201
|||||
701 TTGGGACTGCTGCTCCTCAGCTGTTCATCCCATTTGCTGCGAATTTAGA 750
202 ValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlnGlyIleVala 218
|||||
```

751 GTTGGACCACTGATGTACTGACAGACCTAGAGAGACTAGAGAGGGTGT 800
11thPhProAlaMethIsaMetrPserSerTriPaLProProLeug 235
|||||
801 TACATTTCCAGCCATGCATGCTGTCTTGGGCTCCCTCTTGT 850
235 lnaArgSerLysLeuLeuSerLleSerTyraAglAlaAglInLeuGlyThr 251
|||||
851 AAAGAACCAAACTTTCATGATTCATATGCAGAGACAGCTTGGGACA 900
252 VallleSerLeuProLeuSerGlylleleCysTyrrTyrrMetAnTPth 268
|||||
901 GTAATTTCTCTCTCTCTGGAATATTTGCTACTATATGAATTTGGAC 950
268 rTyTyAlaPheTyrrPhePheGlyThrllleGlyllePheTriPheLeuLeu 285
951 TTATGCTTCTCACTTTTGGTACTATATGGAATTTTGGTCTTCTTGT 1000
285 rPlleTPLeuValSerAspThrProGlnLysHisLysArglleSerHis 301
|||||
1001 GGATCTGTTAGTAGTGACACACCAAAACACAGAGAAATTTCCAT 1050
302 TyrrLulysGluTyrrlleleuSerSerLeuArgAnGlnInleuSerSerG 318
|||||
1051 TATGAAAGGAATACATTTCTTCATATTAAGAAATACGCTTCTTCA 1100
318 nlySerValProTriPaLProlleleuLysSerLeuProleuTriPaL 335
|||||
1101 GAATCTAGTGGCGGTGACCATTTTAAATCCCTCCACTTTGGGCTA 1150
335 lvalValAlaHisPheSerTyrrAnTPthPheTyrrThrLeuLeuThr 351
|||||
1151 TCGAGTTGACACACTTTTCTACAACTGACTTTTATACTTATGTACA 1200
352 leuLeuProThrTyrrMetLysGluilleleuArgPheAnGlnGluAs 368
|||||
1201 TTATGTGCTTCTTATATGAAGAGATCTTAAGGTTCATATGTTCAAGAA 1250
368 nGlyPheLeuSerSerLeuProTyrrLeuGlySerTriPheCysMetlle 385
|||||
1251 TGGGTTTTTATCTTCACTTGGCTTATAGGCTCTTGGTATATGTATCC 1300
385 euSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTriPaSnPheSerThr 401
|||||
1301 TGCTGTGCAAGCTGTGACAAATTAAGGCAAAATATTTTCAACT 1350
402 leuCysValArgArgllePheSerLeuileGlyMetlleGlyProAlaVa 418
|||||
1351 TTATGTGTGCGAGAATTTTATAGCTTATAGGAATGATGACCTGCAGT 1400
418 lPheLeuValAlaAlaGlyPheileGlyCysAspTyrrSerLeuAlaVa 435
|||||
1401 ATTCTGTGCTGCTGCTGCTTCATGCTGTGATATCTTTGGCCGTTG 1450
435 lAphleuThrllleSerThrThrleuGlyGlyPheCysSerSerGlyPhe 451
|||||
1451 CTTTCTCAACTATATCAACAACACTGGGAGCTTTTCTCTCGATGATT 1500
452 SerlleSnHisLeuAspIleAlaProSerTyrrAlaGlylleleuLeuG 468
|||||
1501 AGCATTCACACATCTGGATATGCTCTCTGATGCTGGTATCCCTCGGG 1550
468 ylleThrAspThrPheAlaThrllleProGlyMetValGlyProValllea 485
|||||
1551 CATTACAAATACATTTGCCACTATTCAGAGATGTTGGGCGCTCATTTG 1600
485 lalysSerLeuThrProAspAsnThrValGlyLutrrPginThrValPhe 501
|||||
1601 CTAAGAACTGACCCCGATACACTGTGAGAAATGCAAAACGCTGTC 1650
502 TyrrlleAlaAlaAlaAlaAsnValPheGlyAlaAllePhePheThrLeuPh 518
|||||
1651 TATATGCTGCTGCTATTAATGTTTTTGGTCCATTTTCTTACACTATT 1700

518 eAlaLysGlyLValGlnAsnTriPaLLeuAsnAspHisleGlyHis 535
|||||
1701 CGCCAAAGGTGAGAGTACAAAAGTGGGCTCTCATGATGATCAGACAGACA 1750
535 rGHis 536
|||||
1751 GACAC 1755
seq_name: gb_pr:BC020961
seq_documentation_block: 3329 bp mRNA linear PRI 22-JAN-2002
LOCUS BC020961
DEFINITION Homo sapiens, solute carrier family 17 (anion/sugar transporter), member 5, clone MGC:8885 IMAGE:3847279, mRNA, complete cds.
ACCESSION BC020961
VERSION BC020961.1 GI:18088702
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3329)
Strausberg, R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) md@paxll.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 20 Row: 1 Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6912665.
FEATURES
source
Location/Qualifiers
1..3329
/organism="Homo sapiens"
/db_xref="locusID:26503"
/db_xref="taxon:9606"
/clone="MGC:8885 IMAGE:3847279"
/tissue_type="Colon, adenocarcinoma"
/clone_id="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
125..1612
/codon_start=1
/product="solute carrier family 17 (anion/sugar transporter), member 5"
/protein_id="AAH20961.1"
/db_xref="GI:18088703"
/translation="MRSVRDLARNDSSESDRPLPGARAAAPVCCSARYNLAI
LAFGFRIVYALRNLSVALYDWDNSNTLEDNRTSKACEPHSAPPIVHHNQTKKYQ
WDAETQGMIDGSFFYGIITQIPGIVASKIGKMLGFGILGVALVLTLPPIADYGG
VGPLIVALEAGLDEGVTFPMAHMAWSSMAPLDRSKLISYGAQDGVTLISLPLSG
IICYYMWTYVFYFFGTIGIFWFLIMLVSDPDKRHSIYSYEKEYILRPNVOEGFLSS
OKSDWVPIKSLPLMAIVVAHFSYNMTFVLLLLPVMKEILRPNVOEGFLSSILA
VIGSMICMIISSGOADNLRAKWNSFTLCVRPISLIGMIGPAVLVAAGTFCGYSILA
VAPLTISTTLOGFSSGSINHDIDIASYACILIGITNTPTITPTIGMGPVYIAKSLTPD
NTVGEMQVTFEYIAAIVFAGIAIFFTLRAKGEVQWMLNDHGRH"

BASE COUNT

863 a 718 c 727 g 1021 t

ORIGIN

alignment_scores:

Quality: 2770.50 Length: 528
 Ratio: 5.257 Gaps: 1
 Percent Similarity: 99.811 Percent Identity: 99.432

alignment_block:

us-09-776-865-2 x BC020961 ..

Align seg 1/1 to: BC020961 from: 1 to: 3329

```

10 ArpProValGlnPro...AlaArgProGlyGlyPheGlyLeuSerGlyIar 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
26 CGACCCAGCGGCTCCGGCGCGCTCGGCGCGCTCGGCGCTCGGCGCG 75
25 GARGSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGlyAlaM 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 GCGCTCCCTCTCTGCGCAGGTGGCGAGTACACCTGCTCAGCTAGCGGTCA 125
42 eArgSerProValArgAspLeuAlaArgAsnAspGlyGlyLeuSerThr 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 TGAAGTCTCGGTTGAGACCTGGCCGGAACGATGGCGAGAGAGACAG 175
59 AspArgThrProLeuLeuProGlyAlaProArgAlaGluAlaAlaProVa 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
176 GACCGCAGCGCTCTTCTACGCGGCGCGCCACGCGCGCAGCGCTCCAGT 225
75 LysGlySerAlaArgTyrAsnLeuAlaLeuLeuAlaPhePheGlyPhe 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 GTGCTGCTGCTGCTGTTACACACTTACCAATTTTGGCTTTTGGTTTCT 275
92 heLeValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAspMet 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 TCATGTGTGATGCAATGAGTGAATCGAGTGTGGCTTATGATGATATG 325
109 ValAspSerAsnThrThrLeuGluAspAsnArgThrSerLysAlaCysPr 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
326 GTAGATTCAATAATACACTTTAGAGATATATGAACTTCCAAAGCATGTCC 375
125 ogLHisSerAlaProIleValHisHisAsnGlnThrGlyLysIst 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
376 AGACGATTCCTGCTCCATAAAGTTCAATCAATCAAAAGGTTAAGAGT 425
142 YrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTyr 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
426 ACCAATGGATGCAAGAACTCAAGATGATTCGGGTCTCTTTTAT 475
159 GLyTrIleIleThrGlnIleProGlyGlyTyrValAlaSerLysIleG 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
476 GCGTACATCATCACACAGATTCCTGGAAGATATGTCGCAAGCAAAATAG 525
175 yGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeuThrL 192
526 GGGGAAATGCTGCTAGGATTTGGGATCTTGGCATGCTGTCTCTCACCC 575
192 eupheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleValLeu 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
576 TGTTCACCTCCCATTTGCTGAGATTAGAGTTGAGCACCATCATTTGTACTC 625
209 ArgAlaLeuGluGlyLeuGlyGlyGlyValTyrThrPheProAlaMetHisAl 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
626 AGACCACTAGAGAGACTAGAGAGAGGTGTATCATTTCCAGCAGCATGCTC 675
225 aMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSerI 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
676 CATGTGCTCTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTAGCA 725
242 leSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuProLeuSer 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
726 TTTCAATGAGGAGACAGCTTGGACAGTAATTTCTTTCCTCTTCT 775

```

```

259 GlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
776 GGAATTAATTTGCTACTATATGAAATTTGAGCTATGCTTCTTCTTTTGG 825
275 yThrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerAsp 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
826 TACTATTGGAATTAATTTGTTTCTTTGTGATCTGTTAGTAGTGACA 875
292 hrProGlnLysHisLysArgIleSerHisTyrGluLysGluTyrIleLeu 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
876 CACCACAAAACACAGAGAAATTTCCATTATGAAAGAAAGAAATACATCT 925
309 SerSerLeuArgAsnGlnLeuSerSerGlnLysSerValProTrpValPr 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
926 TCATCATTAAGAATAATCAAGCTTTCTTCAAGAGACAGTCCGCTGGATCC 975
325 oIleLeuLysSerLeuProLeuTrpAlaIleValAlaAlaHisPheSer 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
976 CATTTTAAATCCCTGCCACTTTGGGCTATCGTAGTTGCACACTTTCTT 1025
342 YrAsnTrpTrpThrPheTyrThrLeuThrLeuLeuProThrTyrMetLys 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1026 ACAACTGACTTTTATTACTTATTTGACATTATGCTTACTTATATGAG 1075
359 GluIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeuPr 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1076 GAGATCCTTAAGTTCATATGTTCAAGAGATGGGTTTTATTTCTATGCTC 1125
375 oTyrLeuGlySerTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 392
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1126 TTATTTAGGCTCTGTTATGATGATCCTGTCGTCAGAGCTCTGACA 1175
392 snLeuArgAlaLysTrpAsnPheSerThrLeuCysValArgIleIlePhe 408
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1176 ATTTAAGGGCAAAATGGAATTTTCAACTTATATGTGTCGAGAAATTTT 1225
409 SerLeuIleGlyMetIleGlyProAlaValPheLeuValAlaAlaGlyP 425
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1226 AGCTTATAGGAATGATTTGACCTGCAGATATTCCTGGTAGCTGTGGCTT 1275
425 eIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 442
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1276 CATTTGGCTGTGATTAATCTTTGGCGCTTCTTCTTCACTATATCAACA 1325
442 hrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIle 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1326 CACTGGGAGGCTTTTGCTCTTCTGATTTAGCATCAACCATCTGGAATTT 1375
459 AlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaTh 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1376 GCTCCTTGTATGCTGTGATCTCTGCTGGCATCACAATATATTTGGCAC 1425
475 rIleProGlyMetValGlyProValIleAlaLysSerLeuThrProAspA 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1426 TATTCAGAGATGCTTGGCCCGCTCATTTGCTTAAAGCTCAGCCCTGATA 1475
492 snThrValGlyGluTrpGlnThrValPheTyrIleAlaIleAlaIleAsn 508
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1476 ACACGTGTGGAGATGGCAAAACGCTTCTATATTTGCTGCTGCTATTAAT 1525
509 ValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAs 525
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1526 GTTTTGGTGGCATTTTCTTACACTTATGCGCAAAAGTGAAGTACAAAA 1575
525 nTrpAlaLeuAsnAspHisHisGlyHisArgHis 536
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1576 CTGGGCTCTCATATGATCACCATGACACAGACAC 1609

```

seq_name: gb_cm:AF244578

seq_documentation_block:

LOCUS AF244578 2844 bp mRNA linear MAM 06-AUG-2000
 DEFINITION Ovis aries membrane glycoprotein sp55 (sp55) mRNA, complete cds.


```

1178 AGTTTATAGCTTATAGGATGATGACCTGCATATCTCGTGCCG 1227
423 laqlypheiileglycysaspyrserleuAlaValAlaPheLeuThrlle 439
1228 CAGATTATAGGCTGATGATTATCTTGCTGTGGATTCTTAACATA 1277
440 SerThrThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisle 456
1278 TCACAAACCCGGAGAGCTTTGCTCTTCTTGATTTAGCATCAACCATCT 1327
456 uAspIleAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThr 473
1328 GCACATGCTCTGCTGATGCTGATTCCTGCGCATCCACAATACCT 1377
473 heAlaThrIleProGlyMetValGlyProValIleAlaIlySerLeuThr 489
1378 TTGCCACTATTCCTGATGATGATGGCCCATTCGCCAAGCTTACC 1427
490 ProAspAsnThrValGlyGlyIleThrPGLnThrValPheTyrIleAlaIle 506
1428 CCTGAGAACACTATTTGGAGATGGCAAACTGTTTTCGATCGCTGCTGC 1477
506 AlIAsnValPheGlyAlaIlePhePheThrLeuPheAlaIlySclGlyuv 523
1478 TATCAATGATTTGCTGCATTTCTTCACACATTCGCCAAGGTGAAG 1527
523 alGlnAsnThrIleLeuAsnAspHisIleSclGlyHisArgHis 536
1528 TGCAAACTGGGCGCATGATGATCACCAGACACAGAAC 1568
seq_name: gb_pat:AX207626
seq documentation block:
LOCUS AX207626 2844 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 3 from Patent WO0156598.
ACCESSION AX207626
VERSION AX207626.1 GI:15422331
KEYWORDS
SOURCE
ORGANISM Ovis sp.
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Caprinae; Ovis.
1 (bases 1 to 2844)
Hellerqvist, C.G.
Methods for preventing or attenuating pathoangiogenic conditions by
using the gbs-toxin (cm101) receptor as a vaccine
Patent: WO 0156598-A 3 09-AUG-2001;
VANDERBILT UNIVERSITY (US)
Location/Qualifiers
1. 2844
/organism="Ovis sp."
/db_xref="taxon:9939"
84..1571
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAC60359.1"
/db_xref="GI:15422332"
/translacion="MKSPPVSDIAPSDGERSDRTPLLORAPAEAPVCCSARYNIAF
LSFGFVLYSLRVNLSVALVMDVSNNTAKDNRTSYBEAEHSAPIKVLHNOTKKYR
MGAENOGMILGSEFGYILTOJPGGVASRGSLILGIGITATATFTPLADFG
VGAIVLALBELGEGVYRPAHAMSSWAPPLERKILISYAGAOIGTVVSLDPSG
VICYVMNRYVYFRFVIGIIMFLMICVSDTPETHKITTEYERKYLISLKNQISS
OKSVPIPLKSLPLMAIVAHFSTWTFYLLTLPLTYMKEVLRNIDENGFLSAVP
YLGCWLMILSGQADNLRAWNEFTLWVRVSLGIMGAFIWAAGFIQCDVSLA
VAFIISTLGGFCSSGFSINLIDIASYAGILGITTNPATIPGMIGPIIARSILPE
NTIGEMOTVFCIAAINVFGALFETLFAKGEVQWMAISDHQGRN"

```

BASE COUNT 745 a 623 c 594 g 882 t

alignment_scores:

```

Quality: 2338.00 Length: 497
Ratio: 4.901 Gaps: 0
Percent Similarity: 95.976 Percent Identity: 86.117
alignment_block:
US-09-776-865-2 x AX207626
Align seq 1/1 to: AX207626 from: 1 to: 2844
40 GlyValMetArgSerProValAlaArgAspLeuAlaArgAsnAspGlyGlu 56
78 GGCATCATGAAGTCCCGGTTTCGACTTACCCCGGACGCGGAGGA 127
56 uSerThrAspArgThrProLeuLeuProGlyAlaProArgIleGluAla 73
128 GGGCTCGGACCGCACACCGCTCTGACGGCGCCCGGCGGGAACCCG 177
73 IaproValCysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePhe 89
178 CTCACGATGCTGCTGCTGCTGCTTACAACTPAGCATTTTGTCTTTT 227
90 GlyPhePheIleValIleAlaLeuArgValAsnLeuSerValAlaLeu 106
228 GGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 277
106 IAspMetValAspSerAsnThrThrLeuGluAspAsnArgThrSerIly 123
278 GCACATGCTGATTCMAACACACTGCCAAGATATATGAACGCTTACG 327
123 IAcysProGlnHisSerAlaProIleLeuValHisHisAsnGlnThrGly 139
328 AGTGTGACAGACATTCCTGCTCCATTAAGATTCACAAACCAACGGGT 377
140 LysIlyThrGlnThrPaspAlaGluThrGlnGlyThrIleLeuGlySer 156
378 AAAAGATACCGGTGGATGAGCAAACTCAAGATGATTCGGAATCTTT 427
156 ePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSer 173
428 TTTCTATGCTACATCATCACACAAATTCGAGAGATTTGCCAGCA 477
173 yslileglylyMetleuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
478 GAATGGGGGAGACTGTTGCTAGGATTCGGAGTCTTGTACACGTCAT 527
190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeu 206
528 TTCACCCGTTCACTCCCTGCTGACAGATTCGAGCGGAGCCCTTGT 577
206 eValLeuArgAlaLeuGlyGlyLeuGlyGlyValIleThrProAla 223
578 TGCACCTCAGGCTAGAGGCTAGAGAGGCTGCATATCCAGCA 627
223 eThsAlaMetTrpSerSerTrpAlaProProLeuGluArgSerIlyLeu 239
628 TGCATGCCATGTGCTTCATGGGCTCCCTCTTGAAGAAGAGCTT 677
240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeu 256
678 CTGAGTATTCATATGACAGACAACTTGGACAGATTTCTCTTCC 727
256 oLeuSerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyr 273
728 TCTTTCTGAGTAAATTTGCTACTATATGAATTTGACCTTATCTCTT 777
273 hePheGlyThrIleGlyIlePheThrPheLeuLeuTrpIleTrpLeuVal 289
778 TCTTTGGATGTTGGAATCATCTGTTTATTTATGATCTGCTTACTT 827
290 SerAspThrProGlnIlyHisLysArgIleSerHisTyrGlyLysGly 306
828 AGTATACACCAAGAACCTCAAGACATCACTCCATATGAAGAAGACTA 877

```

```

306 rTleuSerSerleuArgsnGlnleuSerSerGlnlyssSerValProT 323
|||||
878 TATCTTCATCATTAATAAATCAGCTCTCTCACAGAAAGTACGGCGT 927
|||||
323 rPValProIleuLyssSerleuProleuTPAlIleValAlaIahIs 339
|||||
928 GGATACCTATGCTGAATCAGCCACCTTTGGCTATTGTCTTCACAT 977
|||||
340 PheSerTyraSnTrpThrPheThrleuLeuThrleuLeuProThrTy 356
|||||
978 TTTCTTCACTGACGACTTTTATCTTTGTTGACCTTATTGCTTACTTA 1027
|||||
356 rMelysGluIleuLeuArgpheaSnValGlnGlnuSnGlyPheLeuSer 373
|||||
1028 CATGAAGAAGAGCTTAAGTTCAATATTCACAGAAATGGGTTTATCTG 1077
|||||
373 eLeuProTyrlenGlySerTrpLeuCySMelIleuSerGlyGlnAla 389
|||||
1078 CAGTCCCTTATTAGTTGGTTGTTATGTATGATCTCTGGGTCACACT 1127
|||||
390 AlaAspAnleuArgAlaIlystrPasnPheSerThrleuCyValArgAr 406
|||||
1128 GCTGCAATTTAAGGCAAGATGAAATTTTCACTCTGTGGGTCGAAG 1177
|||||
406 gLlePheSerleuIleGlyMetIleGlyProAlaValPheleuValAla 423
|||||
1178 AGTTTATAGCTTATAGGATGATGACCTGCATATTCCTGGTTGCCG 1227
|||||
423 lAglyPheIleGlyCySAspTySerleuAlaValAlaPheleuThrIle 439
|||||
1228 CAGGATTTATAGCTGTGATTTATCTTGGCTGTGATCTCTTAACATA 1277
|||||
440 SerThrThrleuGlyGlyPheCySerSerGlyPheSerIleAsnIle 456
|||||
1278 TCACAACACCTTGAGGCTTTTGCTCTTGATTTAGCATCAACCATCT 1327
|||||
456 uAspIleAlaProSerTyraIaglyIleleuLeuGlyIleThrAsnThr 473
|||||
1328 GCACATTCCTCTCGTATGCTGATTTCTCTGGGCATCAACAATACCT 1377
|||||
473 heAlaThrIleProGlyMetValGlyProValIleAlaLysserleuThr 489
|||||
1378 TTGCCCATTCCTCTGAAATGATGGGCCATCATTCGACAGAGCTTAC 1427
|||||
490 ProAspAnThrValGlyIleTrpGlnThrValPheTyrlleAlaIal 506
|||||
1428 CCGTGAACACATTTGAGAAATGCAAACTGTTTCTGCATCGCTGCTG 1477
|||||
506 aIleAsnValPheGlyAlaIlePhePheThrleuPheAlaLyssGlyIle 523
|||||
1478 TATCAATGATTTGGTGGCATTTTCTTCACACTATTCGCCAAAGGTGA 1527
|||||
523 aGlnAsnTrpAlaIleuSnAspHisGlyHisArgHis 536
|||||
1528 TGCAAACTGGGCGCATGATCATCAAGACACAGAAAC 1568
|||||
seq_name: gb_in:AY060776
seq documentation block:
LOCUS AY060776 1841 bp mRNA linear INV 08-NOV-2001
DEFINITION Drosophila melanogaster GH23975 full length cDNA.
ACCESSION AY060776
VERSION AY060776.1 GI:16768209
KEYWORDS FLI CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Mscocorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1841)
Stapleton,M., Broksstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frishe,E., George,R.,
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,

```

```

TITLE
JOURNAL
COMMENT
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription errors of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
location/Qualifiers
1. 1841
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="92E8-92E8"
1. 1841
/gene="CG4288"
/note="Alignment with genomic scaffold AB003730"
/db_xref="FLYBASE:FBgn0038799"
133. 1641
/gene="CG4288"
/note="Longest ORF"
/codon_start=1
/db_xref="FLYBASE:FBgn0038799"
/product="GH23975p"
/protein_id="AAL28324.1"
/db_xref="GI:16768210"
/translacion="MTDLESPKKNKHLQVHONNNEITSEPTLWFRNKRORYVVL
AFGEFNVYSLRVNLSVAIVAMTENTVDADGNSYQODFPDMSKOKLILSFEYG
VITLOFGVIGTIGKGNIVFGIGSTALITLTPMAHSHLEMLFRTIEGPEG
VPRGIIHAWMARSPRLERSRMASTAPAGNYACTVYAMCSEFLAKKYGWESVYVFG
TTGVITWTIMLVFKAGPELDRECSKECDYLOKTIgyGSKVKKHPRAITSPFT
ALWASHSESNWGFYTLTQLPSEFLRDTLNFEDKGTGILSAVPLAMGILLAVSGYIAD
WLOVKGIWTTTOVRNFGAFLAQTFVFMULAAYLLDPWVSYSLLIAGLAFAMSG
FAVNHDIAPDHASVLMGIGNTFATIPGIVSPLLTGYVVTNOTSDEMRIIFISAGIY
LVGCVTYVWYSCSDLOEMAKTPQKROAEKNAQLQDTAGFVNSGAEKLD"
BASE COUNT 392 a 495 c 479 g 475 t
ORIGIN
alignment_scores:
Quality: 992.00 Length: 448
Ratio: 2.892 Gaps: 3
Percent Similarity: 76.562 Percent Identity: 42.411
alignment_block:
us-09-776-865-2 x AY060776 ..
Align seg 1/1 to: AY060776 from: 1 to: 1841
80 ArgTyraSnleuAlaIleleuAlaPhePheGlyPhePheIleValTyAl 96
|||||
244 CGGTACATTTGGTGGCTACGCTTTTTCGATTCCTTCATGCTACTCTC 293
|||||
96 AleuArgValAsnleuSerValAlaIleValAlaIleValAlaIleVal 113
|||||
294 ACTGCACTGATCTTTCGTCGCAATTTGTCATGACACAGAAATCGCA 343
|||||

```


TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 167928)

REFERENCE
AUTHORS

Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.
Unpublished
2 (bases 1 to 167928)
Worley, K.C.
Direct Submission
Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowe, S., Brieleva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, T., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolle, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowe, S., Brieleva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolle, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowe, S., Brieleva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolle, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowe, S., Brieleva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruiz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolle, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowe, S., Brieleva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Day-Carroll, L., Dederich, D.A., Delaney, K.R., Deigard, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Disper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gablis, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A.,
Hernandez, O., Hernandez, O., Hoogson, A., Hogues, M., Holloway, C.,
Hollins, B., Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louisege, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawlin, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S.,
Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pui, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshari, N.,
Stinson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svalek, A., Taber, P., Tamerisa, A., Tamerisa, K.,

alignment_scores:

Quality: 984.50 Length: 486
Ratio: 2.921 Gaps: 8
Percent Similarity: 69.342 Percent Identity: 40.947

alignment_block:

US-09-776-865-2 x AC023685 ..

Align seg 1/1 to: AC023685 from: 1 to: 167928

80 ArgYrAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValIleValIle 96
149716 CGTACATTTTCGGATTCATGCGCTTCCTGGATTCGCGCGTCAACG 149765
96 AleuArgValAsnLeuSerValAlaLeuValAspMetValAsp..... 110
149766 GATCGGGTCAATCTGTCGGTGGCCATTGCGCCATGCTGACCAACG 149815
111SerAsnThrThrLeuGluAspAsnArgThrSerLysAla 123
149816 CAATTCGCGACGACATCATCGGTGATTCGACGACG..... 149856
124 CysProGluHisSerAlaProIleLysValHisHisAsn..... 136
149857 TGTCCA...CTACGCGACCA.....CATCAATAGGTGACGATCC 149894
137GlnThrGlyLysLysTyrGlnTyrPaspAlaGluThrGlnGly 151
149895 CAATCCGACAGAGAGCGCGAGTTGTGTGGAGCGAGCGACGAGGAT 149944
151 rPileLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGly 167
149945 TGTGTCTCGGAGTTTCTTATGCTATGCTACCAACGAGTCCCGGC 149994
168 GlyTyrValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyIle 184
149995 GGACGAGTGGCGGCTATGCTATGCGAGAGAGATCTACGCTATGAGT 150044
184 eleuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeu 201
150045 GTTATATACGCGGCTTTTACGCTTATATACCTATGCTGCGCACCTGGG 150094
201 LyValGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGlyIle 217

150095 ATCTGCGGCTGTGTGCTGTGTCGCACTCCGAGGAGAAATGGCGAGGCG 150144
218 ValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProIle 234
150145 GTTCACCTATCGATGATGACAGCCATCTGTCCTCCACGAGATACCGCGCT 150194
234 uGlnArgSerLysLeuLeuSerLysIleSerTyrAlaGluIleLeuGly 251
150195 GGACAGAGACAAAGTTCCCGCAATCGTCTATGCGGCGTCCAAATTCGGAA 150244
251 hValIleSerLeuProLeuSerLysIleIleCys.....TyrTyr 264
150245 CAGCATTCCTCATGCTCCGTCGCGAGATGCTGCTGCTGCACTTCCTG 150294
265 MetAsnTrpThrTyrValPheThrPhePheGlyThrIleGlyIlePheTr 281
150295 GGTGGCTGGCGCGCGCTTCTATCATCTCTGAGCTGTGGCATTCGTG 150344
281 pPheLeuLeuTrpIleThrPheValSerAspThrProGlnHisLysA 298
150345 GTTCATGCAATGATGATATTGTTGTTACAGCAACCTCATGATCATCCCA 150394
298 rGlySerHisTyrGluLysGluTyrIleLeuSerSerLeuArgAsnGln 314
150395 GGATCTCAGATCGAGCGAGATATATGAAAGAGTCTACAGTTTCAG 150444
314 314
150445 AGGCTAATAATACAGATCTACGCGAGCGCGAGAGAGAGAGAGAGAG 150494
315LeuSerSerGlnYserValProTyrValProI 326
150495 TGAAGTACATCTGCGGCGCGCGCGAGAGAACCATACCTGTCATCCG 150544
326 IeLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSerTyr 342
150545 TGCTACATCCGTCATCTGTCGGCGCATCTGTGACGCAATGCGCGCAG 150594
343 AsnTrpThrPheTyrThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 359
150595 GGATGGCGCTTTCACGACGACTACGAGCTGCGCCATCAATCAAGACCA 150644
359 uIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeuPro 376
150645 CATCTACACTTTCATCATTCAGATGCTCTGCTCAATGCGGTGCGCT 150694
376 YrIleGlySerTrpLeuLysMetIleLeuSerGlyGlnAlaAlaAspAsn 392
150695 ATCTAACCTCTCTGCTGTCGTCGTCGTCCTGCTGCGCGCGAGATTGG 150744
393 LeuArgAlaLysTrpAsnPheSerThrLeuLysValArgTrpGlyIlePhe 409
150745 ATGCTACGACAGCGCTACATATCTGCTGTAACCTGATATGATGTTGAA 150794
409 rLeuIleGlyMetIleGlyProAlaValPheLeuValAlaAlaGlyPheI 426
150795 CACGCTGCGCTCGTGTGTCATCTGCGCTGATGCGCTGATGATCATCT 150844
426 IeGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThr 442
150845 TGGCTCGATGTCGTCGTCGTCACCTTATGTTGCGCGCGCGCTGCT 150894
443 LeuGlyLysPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAl 459
150895 TTCGCGCGCGCGCTATGCTGTCGTCACACAGATGATATACATAGCGCT 150944
459 aProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThr 476
150945 TCCACGATATGACGACACCATGATGATGATGATGATGATGATGATGAT 150994
476 IeProGlyMetValGlyPro...ValIleAlaLysSerLeuThrProAsp 491
II 491

```

150995 TCTGTGGCTTCCGTGCTCCGTATGTCATCGTCTAATCATCATCATCGC 151044
492 AaThValGlyGlnTrpGlnThrValPheTyrlleAlaAlaLeu 508
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
151045 GAGACTCTACACAGTGCATCTGCTCTTCGTGGCGGCGCTTGAA 151094
508 nValPheGlyAlaIlePhePheThrLeuPheAlaTyGlyGluValGlnA 525
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
151095 TATAGCGGCTACTTCATCTACCTGATCTTCGCGACGCCGAGACCAA 151144
525 snTrpAla 527
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
151145 GCTGCTGC 151152
seq_name: gb_in:AC023711
seq_documentation_block:
LOCUS AC023711 174163 bp DNA linear INV 19-DEC-2001
DEFINITION Drosophila melanogaster 3 BAC RP98-6C4 (Rosewell Park Cancer
Institute Drosophila BAC library) complete sequence.
ACCESSION AC023711
VERSION AC023711.3 GI:17933778
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Munry,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Taboz,P., Williamson,A., Homs1,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratilake,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blanchard,K.,
Diaperport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Drepper,H., Emery-Cohen,A., Ferrilera,S., Gary,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Piltman,G.S., Puri,V.,
Schaefer,F., Shen,H., Strong,K., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,Y., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished
2 (bases 1 to 174163)
Worley,K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 174163)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaratilake,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoque,M., Holloway,C.,
Hollins,B., Homs1,F., Howard,S., Huber,J., Hulysk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovar,J.,
Kovar,C., Kralovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,C., Liu,J., Liu,W.,

```

```

TITLE
JOURNAL
COMMENT
Submitted (19-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 19, 2001 this sequence version replaced gi:6997282.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarly (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

```

FEATURES
source
QUALSTAT-REPORT.
Location/Qualifiers
1..174163
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3"
/clone="RP98-6C4"
BASE COUNT 51926 a 36149 c 35965 g 50523 t
ORIGIN

```

```

alignment_scores:
Quality: 984.50 Length: 486

```

Align seg 1/1 to: AC023711 from: 1 to: 174163

[illegible]

121077 TGAAGTGAAGTCTGGCGGGCCGGCGAGGAAACCGATTACCTGTGATCCG :
326 leuLeuYSerLeuProLeuTriPalalIeValAlaHisPheSerTYR 342
121127 TGGCTACATCCGTCACCTCTGTGGGCGCATCTGTTACGCAAGCGGCCAG :
343 AsnTriPhrPheTYRTrIleuLeuThrIleuProThrTYrMetLysGI 359
121177 GGAATGGGCTTCACAGCCAGCATCAACGAGCGCCACCTACATAGAGCAA 121226
359 uileLeuArGpheaSnValGInGluAsnGlyPheLeuSerSerLeuProT 376
121227 CATTCCTACACTTTCATCATCCAGTCGAATGCTTGCTCATGCGGTCCCT 121276
376 yIleuGlySerTrPLeuLysMetIleLeuSerGlyGlnAlaIAspAsn 392
121277 ATCTAACCTCTGTGTGGTGGCATTTGCTGCTCGCGCCTGGCGGATTTGG 121326
393 LeuArGAlaLysTrIPAsnPheserThrIeucYSValAlaArgIlePheSe 409
121327 ATCTACCCAGACCCCTACATATCTCCTGCTGTAACCTGTAATAGTTGGAA 121376
409 rIleuIleGI MetIleGlyProAlaValPheLeuValAlaIleGlyPheI 426
121377 CACGGTGGCTCCGTGGTGGCATCTGGGCGCTGATGGCATATCTATG 121426
426 IeGlyCAsPTrYSerLeuAlaValAlaPheLeuThrIleSerThrTr 442
121427 TGGGCTGCCATTGGGTGTGGGTACACTTATTTGTGGCCGCGCTCT 121476
443 LeuGIyGI PheCysSerSerGIyPheSerIleAsnHisLeuAspIleAl 459
121477 TTCGGCGGCGCCGCTATGATGGCGAACAGATGATGATACATGACCTCAG 121526
459 aProSerTrYAlaGlyIleLeuLeuGIyIleThrAsnThrPheAlaThrI 476
121527 TCACAGCATATGACAGCCACATATGTCATGCCATCACCAATTCGGCGCAAAA 121576
476 IeProGIyMeTValGIyPro...ValIleAlaLysSerIeueuProAsp 491
121577 TCTGTGGCTTCCTGGCTCCGATGTCACTGGCTTATCATCATCATCTCC 121626
492 AsnThrValGIyGInTrPGLnTrValPheTYrIleAlaIleAlaIleAs 508
121627 GAGACCTGTACACAGTGCGCATCTGTCTTGTGGCGGGCGGGCTGTA 121676
508 nValPheGIyAlaIlePhePheThrLeuPheAlaLysGIyGluValGln 525
121677 TAAAGCGGTGAACCTCATCTACCTGATCTTGTCCGACGCGGAGGACAAA 121726
525 snTriPalal 527
121727 GCTGGTGG 121734

seq_name	gb_in:AE003491
seq_documentation	loc:
LOCUS	AE003491 323461 bp DNA linear INV 04-OCT-2000
DEFINITION	Drosophila melanogaster genomic scaffold 142000013386053 section 8 of 30, complete sequence.
ACCESSION	AE003491 AE002593
VERSION	AE003491.2 GI:10728219
KEYWORDS	HTG.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 323461)
AUTHORS	Adams,M.D., Celisiker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amannalks,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., Georgel,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,

Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,R.G., Chapple,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abriil,J.F., Agbayani,A., An,H.J.,
Andrews-Pfankoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,E., Downes,M., Dugan-Kocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischman,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,M.M., Glasser,K., Glodex,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wel,M.H., Idegawa,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kensilson,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacled,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,Y., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,
Venier,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,O.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
TITLE 2 (bases 1 to 323461)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
On Oct 9, 2000 this sequence version replaced gi:7292821.
Location/Qualifiers
1. 323461
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
join(<15214, .15403,15749, .16006,16069, .16312,17722, .18091)
/gene="fine"
/product="CT563"
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
/protein_id="AAFA48215.2"
/gene="fine"
/note="CG4396"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
join(<15214, .15403,15749, .16006,16069, .16312,17722, .18046)
/gene="fine"
/note="fine gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
/protein_id="AAFA48215.2"
/db_xref="GI:10728220"
/translation="MTNAMDLYKNGSANGSVGNSDESRNLIVYLPOTMGOEMRS
LSSIGLESCKLVADKYSGSISGYGVYVAREDAEKAVNTLNGRLONKVIKSYA
RPSSEIKGANIYVSGLPKRLNSQPDLEGFASFGKITIRILCDNISGLSKVGIRF
DORNAERAIQDLNGRTPGIAEPTIVFAPNAPSAAQIAPLPIAVITPAAATR
RLAAGALPSAGRIIRYSPLAGDLIANSITPGNMTGSGWCIFFVYNLAPEETENLWOLFQ

PFGAQSVKVIIRDLOTSKCKGFGFVTMTNYDEAVVAIQLSLNGCYLGNRYLOVSFTNK
TKT"
complement(join(<23860, .24239,24538, .25184,25286, .25377,
25443, .25533,26089, .26389,26921, .>27095))
/gene="CG4395"
/product="CT4121"
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
complement(<23860, .>27095)
/gene="CG4395"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
complement(join(<23860, .24239,24538, .25184,25286, .25377,
25443, .25533,26089, .26389,26921, .27095))
/gene="CG4395"
/note="CG4395 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
/protein_id="AAFA48216.1"
/db_xref="GI:7292823"
/translation="MATTSSDSESONDVASQAOTQDNURIFLKLHYECVFRONT
VTDTPSSLSLGPATYDSDLPENFSPVPRYLENAAHNEGVIDMRVNDLAEKREBLA
ITVSATMTATNOKENRFLCPNLNEDGYLCMPBPAGVILSOYCPDFEGNRRKLAKTC
LENGSWYRHVPNOTWSNYTNCVDEDEFOFINELVYKGLSLALLISIIIFLG
FLSKCTETRIHVHIFASLACQCAVMITMYRLVRSFETIENAPYRCHEKSYENNTA
LOVAQLKTIITITFQIMCIGLHVHYHFMVYFNMFCBGLHLVLYVSYTSS
GNLAIAGTSLHMAVSHVGRCDLPQPLWQDFTSNLSSESSQWPMALRLVROEMS
GNSVOILPPLSLISQVFNKVDIVRWKTVISWSPRIPAIYVFLINVLRIYVRLH
POSQAPADILKRAVARATILVPLGLHFLIPYRPGDGTQIDHRYOMLSVLSLOG
FVSLPFCFANHDVFAIRITLNLKLPISLVPRPPASNGVMTGMAFTPSRELGV"
join(<43393, .43627,43701, .43859,44240, .45178,45248, .46407,
47115, .>47153)
/gene="CG15721"
/product="CT35951"
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
/evidence=not_experimental
/evidence=not_experimental
/evidence=not_experimental
/evidence=not_experimental
/evidence=not_experimental
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
join(<43393, .43627,43701, .43859,44240, .45178,45248, .46407,
47115, .>47153)
/gene="CG15721"
/note="CG15721 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
/protein_id="AAFA48217.1"
/db_xref="GI:7292824"
/translation="MTAFVLSIGFALLAGARWDYGVYYPAPHYEQOOPYDPCPP
SGPESLNCAGARGAPSTPSPCEQVORRALTGIMMETILHREKCTLEVCNCDQY
NPVCKYKDIRNRNRSCEBLQVRCRTGHPKQKQDGPESRTPVQSPKQVONTNPIA
INPYOPNVSRCAGVSGQPGVSGQPGVSGQPGSPQPPVSSFSVPRPVADPIY
OPYQISWENLPLEYAGTSKPYNPYETPTTSTTASTTSTTSTSTTSTTSTTSTT
DKITTPQPGDSETEASTVPPNPRRNKIVNAVIDTPKSTSTTSTTSTTSTTSTT
SVTTPLRKYITITASPOHATTAPRSKSLISINAVOVPEOEAEKTSPTPTTYSFN
SYNSASYTTPPAVTPEDPTKVPDVEIKIKINVOISARETEKADETQSTLIVSTA
APTNETTLOLPEDKVKINAVESETEKIDEETTKANGCEITPVGTSKAKOTETTK
SLDKVKINAVESETEKIDEETHTYPSATVPSSSVYOTPYDKSSVTPKNDVTYQAP
LEKIVKINAVESETEKIDEETHTYPSATVPSSSVYOTPYDKSSVTPKNDVTYQAP
STEAKDEETTKSLDKVKINAVESETEKIDKAPYSSVTPYSSASVYTPYAS
TEAPKNDETTKMALDLVKINKAJOISTEKAVETPTTEPTITTKPTTYPSQVYTSS
SVYKSSSSPDDDEDEESTNAPLNKSIKINAVSDPTIONTTPATPTPPTTGSGF
HCVKEGPETLDADSVAKSKATSEAEIPYIKTESRNIIGTRDQLEH"
complement(join(<47450, .48160,48223, .48387,48516, .48897))
/gene="CG12716"
/product="CT35950"


```
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
complement(<47450..>48897)
/gene="CG12716"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
complement(join(47703..48160,48223..48387,48516..48897))
CDS
/gene="CG12716"
/feature="1"
/codon_start=1
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
/protein_id="AA048218.1"
/db_xref="GI:7292825"
/translacion="NOTOKRYSNSRAFTMILILQVLGILGRVLAOKSPIANMH
MADPHYSDQYAKILGEAGLODADKEHGLPMLOIYDOPGPGVPCATNG
TDSFEYENHCRLEAANKMLFOHGTLELEPEMERCLPNCOTMCTOVERVCALEIG
GAIPQEFANECMEKHECHTKOYLRIHNGPCOTPKSKGKRLRRKKRPLNLSI
SKPATTPKKVYIMLATPAPRSSSTTPRPSTTTTTRITARRPKPTSTSPKRF
QOLMNVANPMVVSQAIIDAYNVNIPDVGHDXGRDMLLALDVFAHSWSCHRALFRA
"
mRNA
complement(join(<76685..76911,77099..>77279))
/gene="CG15719"
/product="CT35948"
/db_xref="FLYBASE:FBan0015719"
/db_xref="FLYBASE:FBgn0030440"
/evidence=not_experimental
complement(<76685..>77279)
/gene="CG15719"
/map="11D1-11D3"
/db_xref="FLYBASE:FBan0015719"
/db_xref="FLYBASE:FBgn0030440"
/evidence=not_experimental
complement(join(76685..76911,77099..77279))
CDS
/gene="CG15719"
alignment_scores:
Quality: 984.50 Length: 486
Ratio: 2.921 Gaps: 8
Percent Similarity: 69.342 Percent Identity: 40.947
alignment_block:
US-09-776-865-2 x AE003491 ..
Align seg 1/1 to: AE003491 from: 1 to: 323461
80 ArgTYrAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTYrAl 96
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205408 COTCAATTTCGGATTCATGGGCTTCTGGGATTCGCCGTGGCTTACGC 205457
96 AleuArYAlaSnLeuSerValAlaLeuValAspMetValAsp..... 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205458 GATGGGGTCATCTGTCGGTGCATCTGGCCATGCTGAACCAACG 205507
111 .....SerAsnThrThrLeuGluAspAsnArTYrThrSerIlySala 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205508 CAATTCCGACAGCAACTCATCGGATTCGATACGACGACG..... 205548
124 CysProGluHisSerAlaProIleValIshAsn..... 136
||||| |||||| |||||| |||||| |
205549 TGTCCA...CTACCGGCACCA.....CATCAAAATGGTAGCGATCC 205586
137 .....GlnThrGlyLysLYrGlnTrpAspAlaGluThrGlnGlyT 151
||||| |||||| |||||| |||||| |
205587 CAATCCGACAGAGAGGCGAGTTGTGTGGAGACGACGACGCGGAT 205636
151 rPleLeuGlySerPhePheTYrIleIleThGlnIleProGly 167
||||| |||||| |||||| |||||| |
205637 TGTGCTCGGAGTTTCTTATGCGCTATGCTAACCAAGTCCCGCG 205686
```

```
168 GlyTYrValAlaSerLYsIleGlyLYsMetLeuLeuGlyPheGlyI 184
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205687 GCAGGATGGCCGACCTGATGTGGGAAGAATCTACGGCTATGGAGT 205736
184 eLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAspLeuG 201
||| ||||||:|||||:|||||:|||||:|||||:|||||:|||||:
205737 GTTGATCACGGGGCTTATGACCTTATTAATCCATGTGCTGCCACCTGG 205786
201 lYValGlyProLeuIleValLeuArgAlaLeuGlyLYsGlyGly 217
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205787 ATCTCCCGCTGTGGTCTGTGGCCATCTGAGAGGAATGGCCAGGCG 205836
218 ValTrpPheProAlaMetHisAlaMetTrpSerSerTrpAlaProP 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205837 GTACCTTATCCAGCTATGACAGCCATGCTTGCCCACTGGATACGCCCT 205886
234 uGluArGserLYsLeuLeuSerIleSerTYrAlaGlyAlaGlnLeuGlyT 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205887 GGAGAGGAAGAAGTTCCGCCAATGCTGATGCGGGCTCCAAATATCGGAA 205936
251 hrValIleSerLeuProLeuSerGlyIleIecys.....TYrTYr 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205937 CAGTCATTTCCATGCGCGTGGCCGATGCGTGTGCTGCGTACCTCTCG 205986
265 MetAsnTrpTrpTrpValPheTYrPhePheGlyThrIleGlyIlePheTr 281
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205987 GGTGGCTGGCCGCTGGCTTTTCTACATCTTCGACACTGGGCTATTCGTG 206036
281 pPheLeuLeuTrpIleTrpLeuValSerAspTrpProGlnLYsIlyS 298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206037 GTTCATCCCATGATGATGATTTGGTGTACAGCAAGCCTAGCATCATCCA 206086
298 rGlyIleSerHisTYrGlnLYsGlyTYrIleLeuSerSerLeuArGAsnGln 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206087 GGATCTCAAAATCGGAGCGAGATATATCGAAAGAGTCTACAGGTTAG 206136
314 ..... 314
206137 AGGCTAATAAATCAAGATCTAGCGGAGCGAGAGAGAGGAGCAAGCA 206186
315 .....LeuSerSerGlnLYsSerValProTrpValProI 326
206187 TGAAGTGAAGTGGCGGGCGCGCGAGGAACCAATACCTCGTCAATCGC 206236
326 lLeuLYsSerLeuProLeuTrpAlaIleValAlaAlaHisPheSerTYr 342
||||| ||||||:|||||:|||||:|||||:|||||:|||||:|||||:
206237 TGTCTACATCCGTACTCTGTGGCCATCTGTTGACCAATGCGGCCAG 206286
343 AsnTrpTrpPheTYrThrLeuLeuThrLeuLeuProTrpTYrMetLYsG 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206287 GGAATGGCGCTTCTACAGCAGCACTAACCGAGCTGCCCACTACATGAGCA 206336
359 uIleLeuArGpHeAsnValGlnLYsGlyPheLeuSerSerLeuProT 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206337 CATCTTACACTTTCATCCAGTCGAATGCTCTCTCAATCGGTCGCT 206386
376 YrLeuGlySerTrpLeuLYsMetIleLeuSerGlyGlnAlaIleAspAsn 392
||||| ||||||:|||||:|||||:|||||:|||||:|||||:|||||:
206387 ATCAAACTCTGCTGTGCTGGCATTTGCTTCTCGCCGTCGCGGATGG 206436
393 LeuArGAlaLYsTrpAsnPheSerThrLeuLYsValArGArGlyIlePheS 409
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206437 ATGCTAGCCACAAGCTACATATCCCTGTCTGAACTCGTAATAGTTGTGAA 206486
409 rLeuIleGlyMetIleGlyProAlaValPheLeuValAlaIleGlyPheI 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206487 CACGGTGGCTTCGCGTGGTCACTACGCGCTGATGGCATCAATCATATG 206536
426 lGlyCYsAspTYrSerLeuAlaValAlaPheLeuThrIleSerThrTh 442
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206537 TGGGCTGGATTTGGTGTGGTCACTTATGTGGCGCGGCTGGGCTCT 206586
```



```

443 LeuGIyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAl 459
      :::::||||| :::::||||| :::::||||| :::::|||||
206587 TTGGGGGGCCGCTCATCTGTCACACAGATGAATCATACATCCGCTCAG 206636
      :::::||||| :::::||||| :::::||||| :::::|||||
459 aProSerTyrAlaGlyIleLeuLeuGlyIlePheAsnThrPheAlaThrI 476
      :::::||||| :::::||||| :::::||||| :::::|||||
206637 TCCACATATGACAGGACCATGATGACATCACCAATTCGGCGCAATA 206686
      :::::||||| :::::||||| :::::||||| :::::|||||
476 IaProGlyMetValGlyPro...ValIleAlaIleSerLeuThrProAsp 491
      :::::||||| :::::||||| :::::||||| :::::|||||
206687 TCTGTGGCTTCCTGGCTCCGTAATGATCATGCTATCATCATATCCG 206736
      :::::||||| :::::||||| :::::||||| :::::|||||
492 AsnThrValGlyIleTyrPheIleThrValPheTyrIleAlaIleAla 508
      :::::||||| :::::||||| :::::||||| :::::|||||
206737 GAGACTCTACACACATGTCATCTGCTTCTGCTGGCGGCGCTTGA 206786
      :::::||||| :::::||||| :::::||||| :::::|||||
508 nValPheGlyAlaIlePhePheThrPheAlaIleGlyIleValGlna 525
      :::::||||| :::::||||| :::::||||| :::::|||||
206787 TATAGCCGGTAACCTTACTACTCTTCCGCCAGCGGAGACGACAA 206836
      :::::||||| :::::||||| :::::||||| :::::|||||
525 snTPaIa 527
      :::::|||||
206837 GCTGCTCG 206844

seq_name: gb_to:AF271235

seq_documentation_block:
LOCUS AF271235 3982 bp mRNA linear ROD 14-JUN-2000
DEFINITION Rattus norvegicus differentiation-associated Na-dependent inorganic
phosphate cotransporter (DNPI) mRNA, complete cds.
ACCESSION AF271235
VERSION AF271235.1 GI:8515880
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3982)
AUTHORS Mashima,H. and Kojima,I.
TITLE Unpublished
JOURNAL 2 (bases 1 to 3982)
AUTHORS Mashima,H. and Kojima,I.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-2000) Cell biology, Gunma University, Institute
for Molecular and Cellular Regulation, 3-39-15, Shoba-machi,
Maebashi, Gunma 371-8512, Japan
FEATURES
Source
1..3982
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..3982
/gene="DNPI"
463..2211
/gene="DNPI"
/codon_start=1
/product="differentiation-associated Na-dependent inorganic
phosphate cotransporter"
/protein_id="AA076223.1"
/db_xref="GI:8515881"
/translation="MESVKORIIAPRGKIGKFNAGKSLGQIYRLEKKKQDNREITELT
EDGKPLEVPEKAPLDCDCTGCTPRRYIIAISMGLGICISGICNGLVAIVDWNNS
TIHKGAVIKAKFNDPEVGMHSGFEGYIIIGDYGYSIRSLAANVFCAAIL
LSTLNMLIPSAARHVGCVFVRIILOGIVGVTYPAHGHMSKAWAPLESRLATTS
FCGSYAGAVIAMPAGIINOTGSSVYVSGMWYMWMLIYESPKHPTID
ERRRIESTIESGNLIGAMERKTPMKRFTSPRYAIIVANCRSMITFLITSP
AYFEVGFETSKVGLSAVPHLWTLIVPIGGIADFLSKQILSTTVYKILNCGG
GVTALSLVVGYSIRGVASISFLVAVGSGFISGFNNWHLIDIASGRYSILNGISN
GVGTLGMCVPIIVGAMTKNSREMOYVFLIALVHGVITFALPASGSKPMDP
EETSEKCGFIHEDELDEMDITQNTINVGTTSYCATSQENSGMWNGMKKEEFO
ESADAVSYKDRDYS"
3953..3958
polyA_signal

```

```

BASE COUNT 1144 a 823 c 884 g 1131 t
ORIGIN
alignment_scores:
Quality: 973.00 Length: 502
Ratio: 2.695 Gaps: 11
Percent Similarity: 71.912 Percent Identity: 41.235
alignment_block:
US-09-776-865-2 x AF271235 ..
Align seg 1/1 to: AF271235 from: 1 to: 3982
45 ProValArgAspLeuAlaArgAsnAspGlyGlnGluSer..... 57
||| ||||| ::::: ||||| ||||| |||||
531 CCTCGACAGATCTA...CAGGCTCTGGAGAGAAAGCAGATACCCAG 577
||| ::::: ||||| ::::: ||||| ::::: |||||
58 .....ThrAspArgThrProLeuLeuProGlyAlaProArgAla 70
::: ||||| ::::: ||||| ::::: |||||
578 AGACATCGAGCTGACAGAGAGAGCGCAGCCCTGGAGTGGCTGAG... 624
::: ||||| ::::: ||||| ::::: |||||
71 GluAlaAlaProValCys...CysSerAla.....ArgTyr 81
::: ||||| ::::: ||||| ::::: |||||
625 AAGAAAGCTCCGCTATGCGACCTGCTGCTGCGCTGCGCGCGCTA 674
||| ::::: ||||| ::::: |||||
81 rsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeu 98
| ::::: ||||| ::::: ||||| ::::: |||||
675 CATCATAGCATCATGAGCGGCTCGCTTCTGTCATCTCTTGGTATCC 724
||| ::::: ||||| ::::: |||||
98 rValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThr 114
||| ::::: ||||| ::::: ||||| ::::: |||||
725 GCTGTAACTCGGCTGTCGCGCATGTGTGACATGTCACAAACACACATC 774
||| ::::: ||||| ::::: ||||| ::::: |||||
115 LeuGluAspAsnArgThrSerTyrAlaCysProGlnHisSerAlaProI 131
::: ||| ::::: ||||| ::::: |||||
775 .....CACCGGGGTGGCAAAAGTTATCAAGAGAAAGCC..... 807
||| ::::: ||||| ::::: |||||
131 eLysValHisHisAsnGlnThrGlyLysTyrGlnTyrPaspAlaGluT 148
||| ::::: ||||| ::::: |||||
808 .....AAGTTAACTGGGACCCGAG 829
||| ::::: ||||| ::::: |||||
148 hrcInGlyTyrPheLeuGlySerPhePheTyrGlyTyrIleIlePheGln 164
||| ||| ||| ||||| ||||| ::::: |||||
830 CTGTGGGATGATTCACGGGCTCTTCTGTGGGCTATATCATACACAGCAG 879
||| ::::: ||||| ::::: |||||
165 IaProGlyGlyTyrValAlaSerTyrIleGlyLysMetLeuLeuG 181
||| ||||| ||||| ::::: |||||
880 ATCCCGGGGATACATCGCATCGCAGCTGCTAACCGGGCTTGTG 929
||| ::::: ||||| ::::: |||||
181 yPheGlyIleLeuGlyThrAlaValLeuThrPheThrProIleAla 198
| ::::: ||||| ::::: |||||
930 GGTGTCATACCTCTTACCTCTACCTCAATGCTGATGCCATCTGCAG 979
||| ::::: ||||| ::::: |||||
198 IaAspLeuGlyAlaGlyProLeuIleValLeuArgAlaLeuGlnGlyLeu 214
||| ::::: ||||| ::::: |||||
980 CCAGATGCATATGATGATGCGTCATCTTGTATGATATGCAAGACAT 1029
||| ::::: ||||| ::::: |||||
215 GlyGlnGlyValThrPheProAlaMetHisAlaMetTyrSerTyrP 231
||| ||||| ||||| ::::: |||||
1030 GTGGAGGGGTCACCTACCCAGCTGTCAAGGATTTGAGCAAGTGGCC 1079
||| ||||| ||||| ::::: |||||
231 aProProLeuGluArgSerTyrLeuSerTyrIleSerTyrAlaGlyAlaG 248
||| ||||| ||||| ::::: |||||
1080 CCTCTTTGGAGAGAGTAGTGGCTACACCTCTCTCTGCTGCTTCT 1129
||| ::::: ||||| ::::: |||||
248 IaLeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyr 264
||| ::::: ||||| ::::: |||||
1130 ATGCTGACAGCATTCGCAATGCCCTTACCTGCTGCTGAGTAC 1179
||| ::::: ||||| ::::: |||||
265 MetAsnTyrThrTyrValPheTyrPhePheGlyThrIleGlyIlePheThr 281
||| ::::: ||||| ::::: |||||

```


[illegible]

```

seq_documentation_block:
LOCUS      AB032435                3946 bp      mRNA      linear      PRI 18-DEC-2001
DEFINITION Homo sapiens DNPI mRNA for differentiation-associated Na-dependent
            inorganic phosphate cotransporter, complete cds.
ACCESSION  AB032435
VERSION    AB032435.1   GI:7328922
KEYWORDS   differentiation-associated Na-dependent inorganic phosphate
            cotransporter.
SOURCE      Homo sapiens
            Homo sapiens brain thalamus cDNA to mRNA.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (stes)
AUTHORS     Althare,Y., Mashima,H., Onda,H., Hisano,S., Kasuya,H., Hori,T.,
            Yamada,S., Tomura,H., Yamada,Y., Inoue,I., Kojima,I. and Takeda,J.
TITLE       Molecular cloning of a novel brain-type Na(+) dependent inorganic
            phosphate cotransporter
JOURNAL     J. Neurochem. 74 (6), 2622-2625 (2000)
MEDLINE     20281869
REFERENCE   2 (bases 1 to 3946)
AUTHORS     Takeda,J. and Onda,H.
TITLE       Direct Submission
JOURNAL     Submitted (15-SEP-1999) Jun Takeda, Gunma University Institute for
            Molecular and Cellular Regulation, Laboratory of Molecular
            Genetics, Department of Cell Biology, 3-39-15, Showa-machi, Maebashi,
            Gunma 371-8512, Japan (E-mail:j.takeda@akagi.sd.gunma-u.ac.jp,
            Tel:81-27-220-8830, Fax:81-27-220-8889)
FEATURES
     source          location/Qualifiers
     1..3946
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /tissue_type="brain thalamus"
     /note="vector:pcgcm3kf(+); sitel,2.ECOR1"
     414..2162
     /gene="DNPI"
     414..2162
     /gene="DNPI"
     /codon_start=1
     /product="differentiation-associated Na-dependent
            inorganic phosphate cotransporter"

```


523 ValGlnAsnTrpAla 527
||| |||||
1917 AAACAACCCCTGGCA 1931

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 ; Search time 53.59 Seconds

(without alignments)
1730.272 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836

Sequence: 1 MAGAMTPRPVQPARPGF.....LFAKGEVQNALNDHGHRRH 536

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organellar:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	2836	100.0	536	4	Q9NRA2	Q9NRA2 homo sapien
2	2621	92.4	495	4	Q9UGH0	Q9UGH0 homo sapien
3	2329	82.1	495	6	Q9WMD1	Q9WMD1 ovis aries
4	996.5	35.1	559	5	Q9VYG7	Q9VYG7 drosophila
5	992	35.0	502	5	Q9VDM0	Q9VDM0 drosophila
6	970	34.2	582	11	Q9J112	Q9J112 rattus norv
7	970	34.2	582	11	Q920B7	Q920B7 mus musculu
8	965.5	34.0	582	4	Q9P208	Q9P208 homo sapien
9	946.5	33.4	586	5	Q23514	Q23514 caenorhabd
10	941	33.2	560	4	Q9P2U7	Q9P2U7 homo sapien
11	936	33.0	560	11	Q62634	Q62634 rattus norv
12	928.5	32.7	529	5	Q9Y7S5	Q9Y7S5 drosophila
13	917	32.3	483	5	Q61369	Q61369 drosophila
14	895	31.6	466	4	Q96LH1	Q96LH1 homo sapien
15	876	30.9	497	4	Q9Y2C5	Q9Y2C5 homo sapien
16	856	30.2	479	5	Q9W4G7	Q9W4G7 drosophila

17	852.5	30.1	496	5	Q9VKC9	Q9VKC9 drosophila
18	838.5	29.6	481	5	Q9VJW8	Q9VJW8 drosophila
19	822.5	29.0	453	5	Q9NKP8	Q9NKP8 drosophila
20	821.5	29.0	512	5	Q9VFX2	Q9VFX2 drosophila
21	814.5	28.7	465	11	Q91WV5	Q91WV5 mus musculu
22	811	28.6	493	5	Q9VR44	Q9VR44 mus musculu
23	808	28.5	563	5	Q9TZN7	Q9TZN7 caenorhabd
24	798.5	28.2	467	4	Q9H531	Q9H531 homo sapien
25	798	28.1	465	5	Q9V905	Q9V905 drosophila
26	797.5	28.1	573	5	Q09932	Q09932 caenorhabd
27	782.5	27.6	497	5	Q9V8M8	Q9V8M8 drosophila
28	762	26.9	475	5	Q9V8N0	Q9V8N0 drosophila
29	762	26.9	491	5	Q9V8N1	Q9V8N1 drosophila
30	762	26.9	560	5	Q9VOC0	Q9VOC0 drosophila
31	760	26.8	529	10	Q9SD14	Q9SD14 oryza sativ
32	756.5	26.7	462	5	Q9V8M9	Q9V8M9 drosophila
33	747	26.3	512	10	Q82390	Q82390 arabidopsis
34	729	25.7	449	5	Q9V334	Q9V334 drosophila
35	699.5	24.7	390	11	Q921B2	Q921B2 mus musculu
36	696.5	24.6	524	5	Q94886	Q94886 drosophila
37	696.5	24.6	524	5	Q9V763	Q9V763 drosophila
38	643	22.7	495	5	Q9W120	Q9W120 drosophila
39	634	22.4	527	5	Q16923	Q16923 caenorhabd
40	626.5	22.1	420	4	Q9H533	Q9H533 homo sapien
41	583.5	20.6	444	5	Q9W121	Q9W121 drosophila
42	582	20.5	468	5	Q9V9J0	Q9V9J0 drosophila
43	569	20.1	792	5	Q9W123	Q9W123 drosophila
44	566.5	20.0	432	10	Q9FKV1	Q9FKV1 arabidopsis
45	562	19.8	413	10	Q23065	Q23065 arabidopsis

ALIGNMENTS

RESULT 1
ID Q9NRA2 PRELIMINARY; PRT: 536 AA.
AC Q9NRA2:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DT MEMBRANE GLYCOPROTEIN HP59.
GN HP59.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu C., Bardhan S., Cetateanu N.D., Lloyd S.R., Yan H., Carter C.E.,
RA Shi E., Venkov C., Yakes M.F., Page D.L., Hellerqvist C.G.;
RT Identification of a novel membrane protein from mammalian cells that
RT interacts with the anti-pathoangiogenic compound CM101."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF244577; AAF97769.1; -
SQ SEQUENCE 536 AA: 58721 MW: 3DDADFC084007724 CRC64;

Query Match 100.0%; Score 2836; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 1e-217;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGAMTPRPVQPARPGFGLSGRSLICQVASTPAHGVMSRPVRLARDGDESTDR 60
DB 1 MAGAMTPRPVQPARPGFGLSGRSLICQVASTPAHGVMSRPVRLARDGDESTDR 60
QY 61 TPLLPAPRAEAPVCCSRYNLAIAIFGFFIVYALRVNLGVALVDWDSMTTLEDNRT 120
DB 61 TPLLPAPRAEAPVCCSRYNLAIAIFGFFIVYALRVNLGVALVDWDSMTTLEDNRT 120
QY 121 SKACPEHSAPIKVNHNQTKYTQWDAETQWILGSEFFYIITQJPGYVASKIGGKML 180
DB 121 SKACPEHSAPIKVNHNQTKYTQWDAETQWILGSEFFYIITQJPGYVASKIGGKML 180

Oy	181	GFLLIGTAVLLLPFPRIADLVGVGLIYLRALEGEGBEVTPRAHMAHSSNAAPLERSKL	240
Dd	181	GGILIGTVALLFPFPRIADLVGVGLIYLRALEGEGBEVTPRAHMAHSSNAAPLERSKL	240
Oy	241	SISYAGADLGIVISLPLSGIICCYMMNMTYVYFEGGTIGIEFWELLIMVMSDTPQKHRS	3000
Dd	241	SISYAGADLGIVISLPLSGIICCYMMNMTYVYFEGGTIGIEFWELLIMVMSDTPQKHRS	3000
Oy	301	HYEKEYILLSLRNQISSOKSVPAWPIILKSLPLMAIVVAHFSYNTFEYTLTLPTYKKEI	3600
Dd	301	HYEKEYILLSLRNQISSOKSVPAWPIILKSLPLMAIVVAHFSYNTFEYTLTLPTYKKEI	3600
Oy	361	LRFNVOENGFLSSLPYLGSMVLCMTLSQADNLPAKKNFTSLCYRRFEFLIGMIGRAVEL	4200
Dd	361	LRFNVOENGFLSSLPYLGSMVLCMTLSQADNLPAKKNFTSLCYRRFEFLIGMIGRAVEL	4200
Oy	421	VAAIGTIGCDYSLAVALFLLSTPLGFCSSGFSINHLDIAPSYAGIILGINTFEATIPGW	4800
Dd	421	VAAIGTIGCDYSLAVALFLLSTPLGFCSSGFSINHLDIAPSYAGIILGINTFEATIPGW	4800
Oy	481	GPVIAKSLTPDNVTGEMOTVFYIAAIVNFGALFEFTLLPAKGEVONMALNDHGHRRH	5360
Dd	481	GPVIAKSLTPDNVTGEMOTVFYIAAIVNFGALFEFTLLPAKGEVONMALNDHGHRRH	5360
RESULT	2		
Oy	09UGHO	PRELIMINARY;	495 AA.
Dd	09UGHO	PRELIMINARY;	495 AA.
AC	09UGHO		
DT	01-MAY-2000 (TREMBlrel. 13, Created)		
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)		
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)		
DE	StALIN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20047778; Pubmed=10581036;		
RA	Verheijen F.W., Verbeek E., Aulá N., Beers C.E.M.T., Havelaar M.C.,		
RA	Joosse M., Peltonen L., Aula P., Galjaard H., Van der Spek P.J.V.D.,		
RA	Marchal G.M.S.;		
RT	"A new gene, encoding an anion transporter, is mutated in static acid		
RT	storage diseases.";		
RL	Nat. Genet. 23:462-465(1999).		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.		
DR	EMBL: AJ387747; CAB62540.1; -.		
DR	InterPro: IPR003662; sub_transporter.		
DR	Pfam: PF00083; sugar_tr; 1.		
DR	Transmembrane.		
SC	SEQUENCE 495 AA; 54639 MW; 5C6C154B3E93A19E CRC64;		

Query Match	92.4%;	Score 2621;	DB 4;	Length 495;
Best Local Similarity	100.0%;	Pred. No. 1.2e-200;		
Matches 495;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	42	MRVYRDLARDGEESYDRPILPGARAREAPVCCSARVNLIILAFEGFYIYALRVN	101
Db	1	MRSPVRLDARDGEESYDRPILPGARAREAPVCCSARVNLIILAFEGFYIYALRVN	60
QY	102	SVALVDMVDSNTLTEDNRTSKACPEHSAPIKVHHNQGRKYQMDAETQGLILSPFYGI	161
Db	61	SVALVDMVDSNTLTEDNRTSKACPEHSAPIKVHHNQGRKYQMDAETQGLILSPFYGI	120
QY	162	ITQIPGGVYASKIGKMLFGILGTAVLTLFTPIADLDGVLIVYRALEGSGEYTP	221
Db	121	ITQIPGGVYASKIGKMLFGILGTAVLTLFTPIADLDGVLIVYRALEGSGEYTP	180
QY	222	AMHAMSSWAPLERSKILISVAGAOIGTVISLPGILCIYVMNNTYVYFPGTIGIFW	281

Db	161	AMHAMSSWAPPLERSKILLSTVYGAOLGIVTISLPISGIIYNNMTYYVFFPGTIGIFW	240
Qy	282	FLMWIVSDTPQKHRIISHYEKEYILSLRNOLSSOKSVPPWPIKLSPLMAIVVAHFS	341
Db	241	FLMTWIVSDTPQKHRIISHYEKEYILSLRNOLSSOKSVPPWPIKLSPLMAIVVAHFS	300
Qy	342	YNNMTYYVFFPGTIGIFWYNNMTYYVFFPGTIGIFWYNNMTYYVFFPGTIGIFW	401
Db	301	YNNMTYYVFFPGTIGIFWYNNMTYYVFFPGTIGIFWYNNMTYYVFFPGTIGIFW	360
Qy	402	LCVRIEFLSLGIMIGPRAVEFLVAGFICGDYSLAVAFILITSTLIGFCSSGSIINHLDIAP	461
Db	361	LCVRIEFLSLGIMIGPRAVEFLVAGFICGDYSLAVAFILITSTLIGFCSSGSIINHLDIAP	420
Qy	462	YAGILIGITNTFPAIPGAVGVIAKSLTPDNTVGEWQTVFYIAAAINVFCALFFTLPAKG	521
Db	421	YAGILIGITNTFPAIPGAVGVIAKSLTPDNTVGEWQTVFYIAAAINVFCALFFTLPAKG	480
Qy	522	EYONNALNDHHGHRH 536	
Db	481	EYONNALNDHHGHRH 495	

Q9MZD1	3	PRELIMINARY;	PRT;	495 AA.
ID	Q9MZD1			
AC	Q9MZD1;			
DT	01-OCT-2000 (TRMBLrel. 15, Created)			
DT	01-OCT-2000 (TRMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TRMBLrel. 15, Last annotation update)			
DE	MEMBRANE GLYCOPROTEIN SP55.			
CN	SP55.			
OS	Ovis aries (Sheep).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;			
OC	Bovidae; Caprinae; Ovis.			
OX	NCBI_TaxID:9940;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LUNG;			
RA	Fu C., Barthan S., Cetateanu N.D., Lloyd S.R., Carter C.E.,			
RA	Shi E., Venkov C., Yakes M.F., Page D.L., H C.G.,			
RT	"Identification of a novel membrane protein from mammalian cells that			
RT	interact with the anti-pathogenic compound CM101."			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF24578; AAF97770.1; -			
DR	SEQUENCE			
DR	54536 MW; 649D7CA459B28272 CRC64;			

Query Match	82.1%;	Score 2329;	DB 6;	Length 495;
Best Local Similarity	86.3%;	Pred. No. 2.3e-177;		
Matches 427;	Conservative 39;	Mismatches 29;	Indels 0;	Gaps 0;

[illegible]

Qy	342	YNMFFYLLTLTPPYMEILRFVNOENGFLSSLPYIGSMLCMILSQADNINRAWMNST	401
Db	301	YNMFFYLLTLTPPYMEILRFVNOENGFLSSLPYIGSMLCMILSQADNINRAWMNST	360
Qy	402	LCVRRIFSLIGMIGPAFELVAAGFIGDYSLAVAFLLTSTLLGGFCSSGFSINHHDIAPS	461
Db	361	LMVRRVSLIGMIGPAFELVAAGFIGDYSLAVAFLLTSTLLGGFCSSGFSINHHDIAPS	420
Qy	462	YAGILLGITNTFATIPGMGPVYAKSLSTPDNTYGENQYFYTAAINNGAIFFLFAKG	521
Db	421	YAGILLGITNTFATIPGMGPVYAKSLSTPDNTYGENQYFYTAAINNGAIFFLFAKG	480
Qy	522	EVONWALNDHHGRH	536
Db	481	EVONWALNDHHGRH	495
RESULT	4		
09YVG7			
AC	09YVG7	PRELIMINARY;	PRT; 559 AA.
DC	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	CG4330 PROTEIN.		
GN	CG4330.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_Taxid=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY.		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	Georgie R.G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abbil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,		
RA	Foster C., Gabriellian A.E., Gang N.S., Gealbal W.M., Glasser K.,		
RA	Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,		
RA	Hosltin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,		
RA	Jostali M., Kalush F., Kappen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy C., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,		
RA	Palazzo M., Pittman K.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Xe J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster.";		

Query Match	35.1%	Score 996.5	DB 5	Length 559
Best Local Similarity	39.8%	Pred. No. 3.7e-71		
Matches 208	Conservative 91	Mismatches 171	Indels 53	Gaps 10
QY	51	RNDGEESTDR-----TPLLPGAPRAAARVC-CSARYNALIAAFGFFIVALARVLS	102	
DB	31	RSDDEADDEREAFCSGERPRLINSGAABENHCCPKTHHFGFMGLFPAVYAARVLS	90	
QY	103	VALVMDV-----SNTLTEDNRKSKACREHAPLKYHN-----QTKKKYQMDAEPQG	150	
DB	91	VAIVAMVNOTAIPHSNSSVVIDT---CP-LPAP---HHNSDPRPKQKEGFEVDEATQG	143	
QY	151	WLTSEFFGYITQIPRGVYASKTGCKMLLFGILGTAVLFLPFIADLDGVPLIVRA	210	
DB	144	LVLSSEFFGYVLTQVPGGRMAELYGKKIYGVGLITAVFTLTPPLAAMDPLIVRI	203	
QY	211	LEGDEGYTPRPMHMMSSMAPLERSKLSTSYGAOLGVISPLSGIIC---YUNMW	267	
DB	204	LEGMEGYTPRPMHMLAHMIPLERKFAAVYAGSNGIYISMPLAGWLSLDFLGW	263	
QY	268	TYVFFYFEGTIGIFWFLMIWLVSPTQPKHKRISHKEYILSSLRNQ-----	314	
DB	264	PSAFYIFGLIGLIMPIAMVLYVDKPSDHPRISEEREYERSLOVRLINDLAEDEE	323	
QY	315	-----LSQKSVPRVPLIKSLPRLVAIYVHNSYMTTYTLTLPTMYKETLRNV	365	
DB	324	EGODEVSLRAPPEEPIPMSSILTSYPLMALILLTQCGQMAFYTOTELTPYMSNLHMDI	383	
QY	366	OENGFSLSPITGSLWMLILSGAADNLRAKWNFTLCVRBFSLIGMIGAVFLVAGF	425	
DB	384	OSNALLMANVPLITSFVFIACALADMKLARYISLLMSYKIMNVASVVSBLIGITTY	443	
QY	426	IGCDYSLAVAFLLTSTLGGFCSSGFSINHLDIAPSVAGILGTINPTATPGWGP-VI	484	
DB	444	VGCDVWVWTFMLAGVSGRGAIVYACNQNMHIALSPRYAGTIGTNSAANGCFLAPYVI	503	
QY	485	AKSLTPDNTVGEWQTVETIAAIVNEGAIFFTLFAKGEVQWMA	527	
DB	504	GLIINHRETLTQMLHVFWMLAAGLAINAGNFILFIASFASQESWS	546	
RESULT	5			
QYVDMO	PRELIMINARY	PRT	502 AA	
AC	QYVDMO			
DT	01-MAY-2000 (Tremblrel. 13. Created)			
DT	01-MAY-2000 (Tremblrel. 13. Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19. Last annotation update)			
DE	CG4288 PROTEIN (GH23975P).			
GN	CG4288.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:			
OC	Pterygota: Neoptera: Endopterygota, Diptera: Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_taxid=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY.			
RA	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Cocayne J.D.,			
RA	Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.E., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dierckx S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegama C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., Meleod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Y, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AEO03730; AAF55770.1; -;
DR EMBL: AY060776; AA28324.1; -;
DR Flybase; FBgn0038799; CG4288.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KM Transmembrane.
SQ SEQUENCE 502 AA; 56112 MW; 460D6FD1DE741CE2 CRC64;

Query Match 35.0%; Score 992; DB 5; Length 502;
Best Local Similarity 42.4%; Pred. No. 7.4e-11;
Matches 190; Conservative 93; Mismatches 149; Indels 16; Gaps 3;

QY 80 RYNALIAAFGCFIVYALRNLSALVADWDNTLEDNRTSKACRPHSAPIKVNHOTG 139
DB 38 RYIVVLLAFAFGFNVYSLRNLSAIVAMTENTRTVD-----ADGNVSXO-- 83
QY 140 KRYOMDAETGWMILGFFGYITQIPGVYASKGMLGFGILGTAVLTPTPLAAD 199
DB 84 -DEPMDSKQGLIISFFGYITQIPGVYASKGMLGFGILGTAVLTPTPLAAD 199
QY 200 LGVGPLVIALBGLGEGVTPFPAHAMWSSWAPLERSKLLISYAGQIGTVISPLSG 239
DB 143 HSLMEFLVIRIEGFEVGPFIHAWMSPLEERSRASIATAFAGNAGTVAMPSCG 202
QY 260 ITCYMMVTVVFYFEGTIGTFWFLMTWLVSDTPQKHKRIKSHYEKEITLSSLRNQSOK 319

DB 203 FLATKYGMSVFYFGTIGYIWMVLVFKAGBELDFCSKECDYIOKTIGVSGSKHV 262
QY 320 SVPMVPLIKSLPMATVAVHFSYNMWFTYTLTLPTMKILRPNOENGFLSLPLGS 379
DB 263 KHPRAIFTFMPFATVASHFSEMGFTYTLTLPTMKILRPNOENGFLSLPLGS 379
QY 380 WLCMIISGQAADNLRANKNSTLCVRRIFSLIGMIGPAVLVAGFICDYSLAVAFLLT 439
DB 323 GILLAVSGYADWLQYGVITTTQVRRNFGCAFQVTFMMLTAYL-LDPTMSVSLTI 381
QY 440 STLGGRCSSGFSTNHDIAPSYAGILGTITNPFATIPGAVGVIAKSLPDMTVGEMOT 499
DB 382 AVGIGAFWGSFVFNHDLIAPQAHASVLMGNTGTFATIPGIVSPLTLGYVVTNOTSDWMRI 441
QY 500 VFYIAAIVNGAIFFTLFAKGEVOMA 527
DB 442 IFTLSAGIYVGVYIYFCSGDLQEMA 469

RESULT 6
ID 09J112 PRELIMINARY; PRT; 582 AA.
AC 09J112;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE DIFFERENTIATION-ASSOCIATED NA-DEPENDENT INORGANIC PHOSPHATE
DE COTRANSPORTER.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Mashima H., Kojima I.;
RT "Rat DNP1.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF271235; AAF76223.1; -;
SQ SEQUENCE 582 AA; 64575 MW; 99A14F62E685B9E9 CRC64;

Query Match 34.2%; Score 970; DB 11; Length 582;
Best Local Similarity 40.9%; Pred. No. 5e-69;
Matches 201; Conservative 93; Mismatches 165; Indels 32; Gaps 9;

QY 51 RNDGSESTDTPTPLPAP---RAEAPVC-CSA-----RYNALIAFFGFIYALRVNL 101
DB 34 KQNRRTETLLE--DGKPLEVPEKKAPLDCCTGFLPRRTIYAIMSGLCISFGIRCNL 91
QY 102 SVALLVMDVSDNTLEDNRTSKACRPHSAPIKVNHOTGKKYORDAETGWMILGFFGYI 161
DB 92 GVALVDMVNNSTI---HRGKVIKEKA-----KFNMDPEYGMHRSFFWGYI 136
QY 162 ITQIPGVYASKGMLGFGILGTAVLTPTPLAADLGVPGLVLRALBGLGEGVTP 221
DB 137 ITQIPGVYASKGMLGFGILGTAVLTPTPLAADLGVPGLVLRALBGLGEGVTP 221
QY 222 AMAMWSSWAPLERSKLLISYAGQIGTVISPLSGIICYYNMWTVYFPGTIGTM 281
DB 197 ACHGISKMAKAPLERSKLLISYAGQIGTVISPLSGIICYYNMWTVYFPGTIGTM 281
QY 282 FLIMLVSDTPQKHKRIKSHYEKEITLSSLRNLSOK--SVPMVPLIKSLPMATV 316
DB 257 YMWLVSYSPAKHPTIDEERRIIEESIGESANLGMERKRTKRWKFFTSMPYAT 316
QY 337 VAFSTNMWFTYTLTLPTMKILRPNOENGFLSLPLGSGWLMISGQAADNLRANK 396
DB 317 VANFCSWTFYLLISQPAVFEVEFGEISKVMGLSAVPLVWTIIVPIGGQIADLRSK 376
QY 397 WNFSTLCVRRIFSLIGMIGPAVLVAGFICDYSLAVAFITISTLTGGRCSSGFSTNH 456
DB 377 QILSTTVRKINMGCGFMDATLLVVGY-SHTRGVAISPLVAGFSGFAISGFVNH 435

QY 457 DIAPSYAGILGTTTATPGWNGVIAKSLTPDNTVGEWOTVFYIAAINVEGAIFET 516
 DB 436 DIAPRYASILMGISNGVTLSCWPCIIYVGAMTKNKSREEMQYVFLIALVHGVIYFA 495
 QY 517 LFAKGEVQWMA 527
 DB 496 LFASGEKOPMA 506

RESULT 7

Q920B7 PRELIMINARY: PRT: 582 AA.
 AC Q920B7:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE VESICULAR GLUTAMATE TRANSPORTER 2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-21453279; PubMed-11432869;
 RA Bai L., Xu H., Collins J.F., Ghislan F.K.;
 RT "Molecular and functional analysis of a novel neuronal vesicular
 RT glutamate transporter."
 RT J. Biol. Chem. 276:36764-36769(2001).
 DR EMBL; AF324864; AAL08941.1; -
 SQ SEQUENCE 582 AA; 64559 MW; 9E7A4F62E685A8AD CRC64;

Query Match 34.2%; Score 970; DB 11; Length 582;
 Best Local Similarity 40.9%; Pred. No. 5e-69; Mismatches 165; Indels 32; Gaps 9;

QY 51 RNDGEESTDRTPLLPGAP---RAEAPVC-CSA-----RYNLALAEFFFTYVALRVNL 101
 DB 34 KQDNRETIETLE--DGKPLEVEPKKAPLDCCTCGFLPRRYITAIAMSGLFCISFGIRNL 91
 QY 102 SYALVDMVDSNTLEDNRTSKACPEHSADIKYHNNQTKKYQMDAETQGMILGSEFYGYI 161
 DB 92 GAIVDMVNSTT---HRGKVYIKEKA-----KFNWDPETVGMHGSFEMGYI 136
 QY 162 ITQIPGGYVASKIGKMLGFGILGTAVALTFTPIAADGVGRPLVLRLEGLGEVTPR 221
 DB 137 ITQIPGGYVASKIGKMLGFGILGTAVALTFTPIAADGVGRPLVLRLEGLGEVTPR 196
 QY 222 ANHAMSSWAPPLERSKLTISYAGAQLTGVLISPLSGIICYYMMNTYVYFEGTIGIFM 281
 DB 197 ACHGIMSKWAPPLERSRLATISFCGSYAGAVIAMPLAGILVYGTGSSVFIYGGSGMW 256
 QY 282 FILMTLVSDTQOKHKRSHYEKYLSSL---RNOLSSOK--SVWPVILKSLPIAIV 336
 DB 257 YKFWLVSTVSPAKHPTITDEERRYIEESIGSANLGMKEFKTPWRKFTSMPIYAI 316
 QY 337 VAHESYNTVFYLLTLPYKKEILRFNVQENGFLSLPYLSWMLSGAADLRK 396
 DB 317 VANFCRSWTFYLLISQRYFEEVGFELISKYGMLSAVHLVMTIIVPILGGQADLRK 376
 QY 397 WNESTLCVRRISLIGMIGPAVFLVAGFICDYSILAVFLITSTTLGFCSSGFSINL 456
 DB 377 QILSTTVYKIKMNGCFGEAEATLLVGY -SHTRGVAISFLVLAVGFSGFATSGFVNHL 435
 QY 457 DIAPSYAGILGTTTATPGWNGVIAKSLTPDNTVGEWOTVFYIAAINVEGAIFET 516
 DB 436 DIAPRYASILMGISNGVTLSCWPCIIYVGAMTKNKSREEMQYVFLIALVHGVIYFA 495
 QY 517 LFAKGEVQWMA 527
 DB 496 LFASGEKOPMA 506

RESULT 8
 Q9P208 PRELIMINARY: PRT: 582 AA.
 AC Q9P208:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
 DE DIFFERENTIATION-ASSOCIATED NA-DEPENDENT INORGANIC PHOSPHATE
 DE TRANSPORTER.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN THALAMUS;
 RA Alhara Y., Mashima H., Onda H., Hisano S., Kasuya H., Horl T.,
 RA Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
 RT "Molecular cloning of a novel brain-type Na⁺-dependent inorganic
 RT phosphate cotransporter."
 RT J. Neurochem. 0:0-0(2000).
 DR EMBL; AB032435; BAA92874.1; -
 SQ SEQUENCE 582 AA; 64392 MW; CE761E56FA18C6AD CRC64;

Query Match 34.0%; Score 965.5; DB 4; Length 582;
 Best Local Similarity 41.8%; Pred. No. 1.2e-68; Mismatches 154; Indels 27; Gaps 7;

QY 73 APVC-CSA-----RYNLALAEFFFTYVALRVNLVVALVDMVDSNTLEDNRTSKACE 126
 DB 57 APICDCTCGLPRTYITAIAMSGLFCISFGIRNLQVAVDMVNSTT---HNGGVYI 113
 QY 127 HSAPIVHNNQTKKYQMDAETQGMILGSEFYGYIITQIPGGYVASKIGKMLGFGILG 186
 DB 114 KA-----KFNWDPETVGMHGSFEMGYITQIPGGYVASKIGKMLGFGILG 161
 QY 187 TAVLTFTPIAADGVGRPLVLRLEGLGEVTPRANHAMSSWAPPLERSKLTISYAG 246
 DB 162 TSTLNLIPSAARHVGCVIFVRILOGLVGYVPACHGIMSKWAPPLERSRLATISFCG 221
 QY 247 AOLGTYISPLSGIICYYMMNTYVYFEGTIGIFMLTMIWSDPQOKHRSHEKEY 306
 DB 222 STAGAVIAMPLACILVYGTGSSVFIYGGSGMWITMFLIVSYESPANHPTITDEERY 281
 QY 307 ILSSL---RNOLSSOK--SVWPVILKSLPIAIVAHFSYNTVFYLLTLPYKKEIL 361
 DB 282 IEESIGSANLGMKEFKTPWRKFTSMPIYAIIVANFCRSWTFYLLISQRYFEEV 341
 QY 362 RFNVQENGFLSLPYLSWMLSGAADNLRANKNFSTLCYRRIFSLIGMIGPAVFLV 421
 DB 342 GEISKVGMLSAVHLVMTIIVPILGGQADFLRSKQILSTTVYKIKMNGCFGEATLL 401
 QY 422 AAGFICDYSILAVFLITSTTLGFCSSGFSINHDIDIASYAGILGTTTATPGWNG 481
 DB 402 VVGY -SHTRGVAISFLVLAVGFSGFALISGFNVNHDIDIASYAGILGISNGVTLSCGMVC 460
 QY 482 PYANSILTPDNTVGEWOTVFYIAAINVEGAIFETLFAKGEVQWMA 527
 DB 461 PIIVGAMTKNKSREEMQYVFLIALVHGVIYFAIFASGEKOPMA 506

RESULT 9

Q23514 PRELIMINARY: PRT: 586 AA.
 AC Q23514:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-2001 (TREMBLrel. 17, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE HYPOTHETICAL 65.0 KDA PROTEIN.
 GN HK54.1.

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=9069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.; Gattung S.; Lee T.T.;
RL "The sequence of C. elegans cosmid ZK54.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58737; AAK39396.1; -.
KW Hypothetical protein.
SQ SEQUENCE 586 AA; 65045 MW; 95D52B7F7FA7DD0F CRC64;

Query Match 33.4%; Score 946.5; DB 5; Length 586;
Best Local Similarity 42.6%; Pred. No. 3.8e-67;
Matches 193; Conservative 79; Mismatches 154; Indels 27; Gaps 6;

OY 80 RYMLATIAFGFEFTYVALRNLSVALVMDSTLTEDNTSKACRPHSAPIKVHNQGT 139
DB 145 RHVVAITALLGFANIYAMRNLSIAIVEMT-SETERKVNSTLT-----HVLG 190
OY 140 KKQVDAETQGWILGSFFYYITITQPGVAVASKGKMLGFGILGTAVLTFTPIAD 199
DB 191 DFNMTPTMGOVVLSEFFYITVSQLPGYLATTHGAKTTFAGTGTAVLTTPPFAR 250
OY 200 LGVGPLIVLRALGELGEGVTFRPMAMWSSWADPLERSKLLSTSYAGAQLTYSPLSG 259
DB 251 MGSGMLVFAFMFGMLLEGVYTPAMHVIWMSRMAPMEQTKLATPAFSGSYGTVAAPLSA 310
OY 260 IICYYMMVTVEFEFGTIGTFWFLMTLVSDTPQKHKRISHKEKYLISLNQSSOK 319
DB 311 YLDEHFGWPIEFWFEGALGVIMCWVWKTVHDPEDDPKISTSE---TALLQPDVASON 366
OY 320 S--VPWVPIILKSLPLMAIVVAHSYMTFTLTLLPTMYKEILRFVNOENGFLSLPYL 377
DB 367 HYIVPAQAILRSKRPVNAVIVAHSAQNLGFIYIMLTNLPKMLKDLAGIVVEAGLASSLPYF 426
OY 378 GSWLMLSGQAADNLRAKNWSTLCVRRIISLIGMIGRAVPLVAAFGTICDYS--LAV 434
DB 427 LMGFOIITGGOLDYLRDRDHDTLFRKMACALGFIGGVFL---FLVMTTSNLSLLV 482
OY 435 AFLTISTTLGGPSSGFSINHLDIAPSYAGILGINTPTATPTGAMGVPYAKSLTFDNTY 494
DB 483 LFFSISIGLGGICWCGFSVNHDLAPQYAGHLMATSNFTATPTGIGPLVGAIVONGTI 542
OY 495 GEMQVEYIYAALNVFCAIFETLFAKGEVONMA 527
DB 543 GEMNVIMYITISAVLGAALTKFAADATTDPA 575

RESULT 10
O9P2U7 PRELIMINARY; PRT; 560 AA.
AC O9P2U7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)

DE BRAIN-SPECIFIC NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
GN BNPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Aihara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T.,
Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
RT "Molecular cloning of a novel brain-type Na+-dependent inorganic
phosphate cotransporter.";
RL J. Neurochem. 0:0-0(2000).
DR EMBL; AB032436; BAA92875.1; -.
SQ SEQUENCE 560 AA; 61613 MW; C8BDAPF34B6E45B6 CRC64;

Query Match 33.2%; Score 941; DB 4; Length 560;
Best Local Similarity 39.5%; Pred. No. 9.8e-67;
Matches 193; Conservative 93; Mismatches 162; Indels 40; Gaps 8;

OY 53 DGEESIDRT---PLLP-----GAPRAEAPVCCSARKNLALIAFGFEFTYVALRVNLSVAL 105
DB 38 DGRPVTTQTGRDPVVDCTCFGLPR-----RYIIAIMGIGFCISFGIRCNLGVAI 87
OY 106 VDMVDSNTLTEDNRTSKACRPHSAPIKVHNQGTQWMAEFGWILGSFFYYITITQI 165
DB 88 VSMVNNSTT-----HRGHHVYVQKAQ---FSMPDETVALHGSFFGVIYVTOI 132
OY 166 PGGVAVASKGKMLGFGILGTAVLTFTPIADLGVPLIVLRALGELGEGTFRPMAHA 225
DB 133 PGGFIOCKFANVFPFAIVATSTLNLIPSAKVHNGCVIFRILQGLGEGVTYACHG 192
OY 226 MMSWADPLERSKLLSTSYAGAQLTYSPLSGIICYYMMVTVEFEFGTIGTFWFLW 285
DB 193 IWSKMAPLERSKLLATPAFEGSYAGAVAMPAGVLYOVYSGMSVFFYVSGIFWYLFW 252
OY 286 IWLVSQDPQKHKRISHKEKYL-----LSSLNQSSOKSVWVPIILKSLPLMAIVVAH 339
DB 253 LVAYESPALHPSISEERKYLIDALGESAKLNPL-TKSTYWRPFSTMPYAITIVAN 311
OY 340 FSYNMTFTYTLTLPTMYKEILRFVNOENGFLSLPLSGWMLCMILSGQAADNLRAKNW 399
DB 312 FCRSWTFYLLISQPAVFEVFEFELISKVGLVSAFLPLVFTIIVPGQIADLRSRRIM 371
OY 400 STLTVRRIESLIMIGRAVPLVAAFGTICDYSIAVAFITSTTLGGPSSGFSINHLDA 459
DB 372 STTNVVRKIMNCGFGMEATLLVGVY-SHSKGYAISFLVAVGFSGRAISGFVNHDLIA 430
OY 460 PSYAGILGINTPTATPTGAMGVPYAKSLPDMTVGEMQVEYIYAALNVFCAIFETLFA 519
DB 431 PRYASITIMGISNGVTLGKWCPIIVGAMTKHKTREMOYVPLIASLVHGVIFYGEVA 490
OY 520 KGEVONMA 527
DB 491 SGEKQRPMA 498

RESULT 11
O62634 PRELIMINARY; PRT; 560 AA.
AC O62634;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE BRAIN SPECIFIC NA+-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

Query	Subject	Accession	Score	E-value	Length	Ident	Mismatches	Gaps
Q1	Medicine	94261635	33.04	0.000000	560	105	168	0
Q2	Medicine	94261635	33.04	0.000000	560	105	168	0
Q3	Medicine	94261635	33.04	0.000000	560	105	168	0
Q4	Medicine	94261635	33.04	0.000000	560	105	168	0
Q5	Medicine	94261635	33.04	0.000000	560	105	168	0
Q6	Medicine	94261635	33.04	0.000000	560	105	168	0
Q7	Medicine	94261635	33.04	0.000000	560	105	168	0
Q8	Medicine	94261635	33.04	0.000000	560	105	168	0
Q9	Medicine	94261635	33.04	0.000000	560	105	168	0
Q10	Medicine	94261635	33.04	0.000000	560	105	168	0
Q11	Medicine	94261635	33.04	0.000000	560	105	168	0
Q12	Medicine	94261635	33.04	0.000000	560	105	168	0
Q13	Medicine	94261635	33.04	0.000000	560	105	168	0
Q14	Medicine	94261635	33.04	0.000000	560	105	168	0
Q15	Medicine	94261635	33.04	0.000000	560	105	168	0
Q16	Medicine	94261635	33.04	0.000000	560	105	168	0
Q17	Medicine	94261635	33.04	0.000000	560	105	168	0
Q18	Medicine	94261635	33.04	0.000000	560	105	168	0
Q19	Medicine	94261635	33.04	0.000000	560	105	168	0
Q20	Medicine	94261635	33.04	0.000000	560	105	168	0
Q21	Medicine	94261635	33.04	0.000000	560	105	168	0
Q22	Medicine	94261635	33.04	0.000000	560	105	168	0
Q23	Medicine	94261635	33.04	0.000000	560	105	168	0
Q24	Medicine	94261635	33.04	0.000000	560	105	168	0
Q25	Medicine	94261635	33.04	0.000000	560	105	168	0
Q26	Medicine	94261635	33.04	0.000000	560	105	168	0
Q27	Medicine	94261635	33.04	0.000000	560	105	168	0
Q28	Medicine	94261635	33.04	0.000000	560	105	168	0
Q29	Medicine	94261635	33.04	0.000000	560	105	168	0
Q30	Medicine	94261635	33.04	0.000000	560	105	168	0
Q31	Medicine	94261635	33.04	0.000000	560	105	168	0
Q32	Medicine	94261635	33.04	0.000000	560	105	168	0
Q33	Medicine	94261635	33.04	0.000000	560	105	168	0
Q34	Medicine	94261635	33.04	0.000000	560	105	168	0
Q35	Medicine	94261635	33.04	0.000000	560	105	168	0
Q36	Medicine	94261635	33.04	0.000000	560	105	168	0
Q37	Medicine	94261635	33.04	0.000000	560	105	168	0
Q38	Medicine	94261635	33.04	0.000000	560	105	168	0
Q39	Medicine	94261635	33.04	0.000000	560	105	168	0
Q40	Medicine	94261635	33.04	0.000000	560	105	168	0
Q41	Medicine	94261635	33.04	0.000000	560	105	168	0
Q42	Medicine	94261635	33.0					

RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S.C., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abtill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew K.M., Basu A., Bendale J., Bayraktaroglu D., Beasley E.M.,
RA Beeson K.Y., Besos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fodor C.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fraser C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jalali M., Kalush F., Karpen G.T., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.Z., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel Y., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Myrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavertl J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*";
RA science 287:2185-2195(2000).
[2]
RP SEQUENCE OF 449-529 FROM N.A.
RC STRAIN=CANTON-S;
RA Da Lage J.-L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
CC SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL: AE003806; AAF57968.1; -
DR EMBL: AE003806; AAF57969.1; -
DR EMBL: AF022713; AAD09148.1; -
DR FlyBase: FBgn0024315; Picot.
DR InterPro: IPR003362; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
RW Alternative splicing: Transport; Transmembrane.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 429 449 POTENTIAL.
FT TRANSMEM 466 486 POTENTIAL.
FT VASPLIC 1 35 MPFRSSLNHHRRCHVLYVMQNRNLHELEQQPQR -> MS
SQ ASKEVIGSTGKDEKPAKG (IN SHORT ISOFORM).
SEQUENCE 529 AA; 58372 MW; 07BB9A52D5081EFB CRC64;

Query Match 32.7%; Score 928.5; DB 5; Length 529;
Best Local Similarity 41.1%; Pred. no. 9e-66;
Matches 190; Conservative 81; Mismatches 176; Indels 15; Gaps 8;

76 CCSARYNLAIAAFGEFFTYALRVNLVAALVDMDVSNTTLEDNRKTS--ACPEHSAPIK 132
36 CEATGYETTFMFLGMANAAAYVMRTNNMSYAIVAMV-NHTPAIKSGEAEAYDECGDRIPIT- 93

[illegible]

RESULT	15
09Y2C5	
ID	09Y2C5
AC	PRELIMINARY;
DT	09Y2C5;
DT	01-NOV-1999 (TReMBLrel. 12, Created)
DT	01-NOV-1999 (TReMBLrel. 12, last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE	NA/P04 COTRANSPORTER HOMOLOG.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=99253143; PubMed=1031585;
RA	Shibui A., Tsunoda T., Seki N., Suzuki Y., Sugano K., Sugano S.;
RT	"Isolation and chromosomal mapping of a novel human gene showing
RT	homology to Na ⁺ /PO4 cotransporter.";
RL	J. Hum. Genet. 44:190-192(1999).
CC	-1- SUBCELLULAR LOCATION: INTERAL MEMBRANE PROTEIN (BY SIMILARITY
CC	-1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC	EMBL; AB020527; BAA76663.1; -
DR	Interpro; IPR003662; sub_transporter.
DR	Pfam; PF00083; sugar_tr.1.
KN	Transmembrane.
SO	SEQUENCE 497 AA; 54055 MW; 5A5616AA52D4990D CRC64;

Query Match	30.9%;	Score 876;	DB 4;	Length 497;
Best Local Similarity	37.8%;	Pred. No. 1.3e-61;		
Matches 176;	Conservative 98;	Mismatches 182;	Indels 10;	Gaps 3

```

0Y 77 CSARINLIIAFEGFEIYVALRVNLVSLVADYVDSNTLLEDRFSKXCPREHSAPIKYHN 136
Db 33 CSVRHGLALIILQCLNFSYITIQOMMLSTAIRPMV--NNTAPPSQPNASTERPDSQSYWN 90
0Y 137 QTKRK-----YQMDAETQGMILGSEFFGYIITIDPEGVYASKIGKMLGFEILGTAV 189
Db 91 ETLKEFKMADAVDMSPEITQIILSLNYSGLAPIPSGYAIGFAKYYVAGLFTSSF 150
0Y 190 LTLTFLPIAADLGVGBLYLRALBESLGBGYTFPAMHAMSSAPLERKSLISTISYAGOL 249
Db 151 LTLTFLPIAANAGVALLIYLRIVQIGIAQVMVLGTQYSIVMWAPLERSSOLTTINGSSML 210
0Y 250 GTVLSLPLSGILCYMMNTVYFEFEGHIGLFWELTLMIVSDPQKRRKISHYEKETLS 309
Db 211 GSFYIYLAGGLCOTIGTPYFIYFGGACACPLMPLIYDDVNHPIFSAGKRYIVC 270/0
0Y 310 SLRNO-LSSQKSVMPVPLKLKSLPMATIVVAHFSYNTWTFYTLTLTPRYMEKILEFNQEN 368
Db 271 SLAQDQSPGMSLPIRMKIKSLPMALILVSYFCGYWLEFYTIMATPYIISVLANLRDS 330
0Y 369 GFLSLPYLGSWLCMIISGQADNLRKMNKNSYLCVRIFSLIGMIGCPAVELVAAGTGC 428
Db 331 GILSLALPFWVVCICIIILGGLADLRLKILIRTLITRKLETAISGLVLPSPVILSLPWVRS 390
0Y 429 DYSILAVFLTISTLGGRCSSGSFINHLDIAPSAUGLLGITNFPAITIPGNGVIAXSL 488
Db 391 SHSMTEFLVLSAISSCESGALVNPDLAPRTGGLKGLQVFAHIAQAIPTAAGTF 450

```

```

QY      489  TPDNVTGEMQTVFYIAAAINVEGALFFTLFAKGEVQNMALNDHHGH 534
      ::  |  ::  |  ::  |  ::  |  ::  |  ::  |  ::  |  ::  |
Db      451  ISQDSEFGWRNVFLLSAAVNISGLFYLLFGRADVQDMAKEQTEETH 496

```

```
Search completed: July 15, 2002, 08:25:02
Job time: 167 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 ; Search time 56.81 seconds
(without alignments)
1047.977 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836

Sequence: 1 MAAGAMTPPRVQPARPGF.....LFAKGEVQNALNDHGHHRH 536

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_032802:*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2836	100.0	536	21	AAV45089 Human GBS toxin re
2	2836	100.0	536	22	AAE06518 Human group B beta
3	2621	92.4	495	21	AAV45087 Patial human GBS
4	2621	92.4	495	22	AAAB6967 Human ASF. Homo s
5	2614	92.2	495	22	AAAM38959 Human polypeptide
6	2329	82.1	495	21	AAV45088 Sheep GBS toxin re
7	2329	82.1	495	22	AAE06519 Sheep group B beta
8	1698	59.9	495	21	AAV45090 Human/Sheep consen
9	1679	59.2	314	22	AAAG6538 Human sodium depen
10	1614	56.9	309	22	AAAM40745 Human polypeptide
11	1450	51.1	277	22	AAAM93914 Human polypeptide,

12	996.5	35.1	559	22	ABB58701 Drosophila melanog
13	992	35.0	502	22	ABB60525 Drosophila melanog
14	992	35.0	502	22	ABB65873 Drosophila melanog
15	959.5	33.8	582	22	AAW79273 Human protein S60
16	941	33.2	567	22	AAO13870 Human polypeptide
17	940	33.1	560	17	AAW05148 Human brain sodium
18	940	33.1	560	17	AAW70500 Human sodium-lithi
19	939	33.1	194	22	AAW25685 Human protein sequ
20	928.5	32.7	516	22	ABB67013 Drosophila melanog
21	928.5	32.7	529	22	ABB63684 Drosophila melanog
22	856	30.2	479	22	ABB62841 Drosophila melanog
23	852.5	30.1	496	22	ABB61407 Drosophila melanog
24	838.5	29.6	481	22	ABB64204 Drosophila melanog
25	821.5	29.0	512	22	ABB60925 Drosophila melanog
26	811	28.6	493	22	ABB59580 Drosophila melanog
27	808	28.5	576	20	AAW86523 Eac-4 protein amn
28	798	28.1	465	22	ABB64710 Drosophila melanog
29	782.5	27.6	497	22	ABB70142 Drosophila melanog
30	762	26.9	475	22	ABB67155 Drosophila melanog
31	762	26.9	491	22	ABB70144 Drosophila melanog
32	762	26.9	560	22	ABB59401 Drosophila melanog
33	756.5	26.7	462	22	ABB70143 Drosophila melanog
34	755	26.6	343	22	ABG12999 Novel human diagno
35	747	26.3	512	21	AAAG29915 Arabidopsis thalia
36	747	26.3	519	21	AAAG29914 Arabidopsis thalia
37	731	25.8	395	21	AAAG29916 Arabidopsis thalia
38	729	25.7	449	22	ABB58659 Drosophila melanog
39	708	25.0	256	22	ABG02032 Novel human diagno
40	696.5	24.6	524	22	ABB64935 Drosophila melanog
41	686	24.2	436	19	AAW78919 Human haemochromat
42	653	23.0	425	22	AAAG81307 Human AFP protein
43	643	22.7	194	22	ABB64689 Drosophila melanog
44	615.5	21.7	401	19	AAW79920 Human haemochromat
45	615.5	21.7	401	19	AAW69971 Human sodium-depen

ALIGNMENTS

RESULT 1	
ID	AAV45089 standard; Protein: 536 AA.
XX	
AC	AAV45089;
XX	
DT	31-MAY-2000 (first entry)
XX	
XX	Human GBS toxin receptor (HP59).
XX	
KW	Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW	pathological vascularisation; Cancer metastases; angiogenesis;
KW	neovascularisation; reperfusion injury; scarring; keloid;
KW	chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW	endothelial cell proliferation; antibacterial; anticancer;
KW	anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Modified-site
FT	/note= "Putative amidation site"
FT	138..141
FT	Modified-site
FT	/note= "Putative amidation site"
FT	100..103
FT	Modified-site
FT	/note= "Asn is N-glycosylated"
FT	112..115
FT	Modified-site
FT	/note= "Asn is N-glycosylated"
FT	118..121
FT	Modified-site
FT	/note= "Asn is N-glycosylated"
FT	136..139
FT	Modified-site
FT	/note= "Asn is N-glycosylated"
FT	266..269
FT	Modified-site
FT	/note= "Asn is N-glycosylated"

FT Modified-site 343..346
/note="Asn is N-glycosylated"
FT Modified-site 398..401
/note="Asn is N-glycosylated"
FT Modified-site 297..300
/label="CAMP_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 113..116
/label="CK2_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 114..117
/label="CK2_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 300..303
/label="CK2_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 493..496
/label="CK2_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 66..71
/note="Putative myristylation site"
FT Modified-site 167..172
/note="Putative myristylation site"
FT Modified-site 183..188
/note="Putative myristylation site"
FT Modified-site 213..218
/note="Putative myristylation site"
FT Modified-site 246..251
/note="Putative myristylation site"
FT Modified-site 250..255
/note="Putative myristylation site"
FT Modified-site 378..383
/note="Putative myristylation site"
FT Modified-site 427..432
/note="Putative myristylation site"
FT Modified-site 444..449
/note="Putative myristylation site"
FT Modified-site 464..469
/note="Putative myristylation site"
FT Modified-site 468..473
/note="Putative myristylation site"
FT Modified-site 23..25
/label="PKC_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 58..60
/label="PKC_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 78..80
/label="PKC_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 120..122
/label="PKC_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 138..140
/label="PKC_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 310..312
/label="PKC_phospho_site
/note="Putative phosphorylation site"
FT Modified-site 317..320
/label="PKC_phospho_site
/note="Putative phosphorylation site"
PN MO200005375-A1.
XX 03-FEB-2000.
PD 22-JUL-1999; 99WO-US16676.
PF 22-JUL-1998; 98US-0093843.
PR (UYVA-) UNIV VANDERBILT.
XX

PI Hellerqvist CG, Fu C;
XX WPI: 2000-205377/18.
DR N-PSDB; AA250879.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Claim 10; Page 93-95; 109pp; English.
XX
CC The present sequence is partial human GBS (group B beta-hemolytic
CC streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
CC protein with seven transmembrane domains. Expression vectors comprising
CC the coding region can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularization, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
SQ Sequence 536 AA;

Query Match 100.0%; Score 2836; DB 21; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.3e-283;
Matches 536; Conservative 0; Mismatches 0; Indels 0;

QY 1 MAAGAMPPPVOPARCGFGLGSRSLCOVASTPAPHVGVMSPYVDLARNDEESTDR 60
DB 1 maagampppvoparpgfglsgrrslcqvastpahgvmspyrdlarndeestdr 60
QY 61 TPLPGAPRAEAAPVCCSARYNLAIAFEGFTIVYALRVNLVYALVDMVDSNTLEDNRT 120
DB 61 tplpgapraeaapvccsarynlaiafegftivyalrvnlvsvalvdmvdsntlednrt 120
QY 121 SKACPEHSAPIKVHHNQTGRKRYOMDAETQGWILGSPFYGIITQIRGGYASXIGSKML 180
DB 121 skacpehsapikvhhnqtgrkryqwdetqgwilgspfygiltqirpgyvasklqgkml 180
QY 181 GFGILGRAVLTLPPIADAGVPLIVLRALFEGEGVPPRAPHAMWSSAPPLERSKLL 240
DB 181 gfgilgravltlppiaadagvplilvrlalfegegvpvrpaphamwssvapplerskll 240
QY 241 S1SYAGAQLGTVISLPLSGIICYYMMNTVYFEGFTIGIFWFLMLTWLVSDFPQKHRS 300
DB 241 s1syagaqlgtvislplsgiiCYMMNTVYFfgftigifwflmltwlvsdtpqkhris 300
QY 301 HYKEYILSLRNOLSSQKSVPPVPIIKSLPMAIVVAHRSYMMWTFYTLTLPTWKET 360
DB 301 hykeyilslrnolssqkswppvpilkslpmaiVVAHRSYMMWTFYTLTLPTWkET 360
QY 361 IRRNVOENGLSSLPYIGSWLMTLSGOADNLRANKNFSTLCVRRFRSLIGMIGAVFL 420
DB 361 lrrnvoenGLSSLPYIGSWLMTLSGOADNLRANKnfstlcVRRFRSLIGMIGAVFL 420
QY 421 VAAGFIGCDYSLAVAFLLTSTLGGFCSSGFSINHLDIASVAGILGTTNFTATIPGMV 480
DB 421 vaagfigcdyslavaflltstlGGFCSSGFSinhldiasvAGILGTTNFTATIPgmV 480
QY 481 GPVIAKSLTPDNTVGEWQYFYTAALINVGAFITFLPAKGEVQNNALNDHRRH 536
DB 481 gpviaksltpdntvgewqfytaaalinvGAFITFLPAKGEVqNNALndhrrh 536
RESULT 2
AAE06518
ID AAE06518 standard; Protein; 536 AA.
XX

AC	AAE06518;	
XX		
DT	16-OCT-2001	(first entry)
DE		
XX		
XX		
KW	Human group B beta-haemolytic Streptococci toxin receptor (HP59) protein.	
KW	Human: group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;	
KW	cytostatic; vulnerary; antiatherosclerotic; osteopathic; vasotropic;	
KW	prevention; attenuation; pathoangiogenic condition; cancer; scar;	
KW	wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;	
KW	keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;	
XX	vaccine.	
OS		
XX	Homo sapiens.	
Key	Location/Qualifiers	
FH	8..28	
FT	/label= Hab3_immunogenic_peptide	
FT	Region	
FT	49..63	
FT	/label= Hab1_immunogenic_peptide	
FT	49..76	
FT	/label= Hab4_immunogenic_peptide	
FT	112..125	
FT	Region	
FT	/label= Hab2_immunogenic_peptide	
XX		
PN	WO200156598-A2.	
XX		
PD	09-AUG-2001.	
XX		
PE	02-FEB-2001; 2001WO-US03662.	
XX		
PR	02-FEB-2000; 2000US-0179870.	
XX		
PA	(UYVA-) UNIV VANDERBILT.	
XX		
PI	Hellerqvist CG;	
XX		
DR	WPI: 2001-488844/53.	
XX		
DR	N-PSDB: AAD10325.	
XX		
PT	Preventing or attenuating pathoangiogenic conditions e.g. cancer,	
PT	chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by	
PT	administering group B beta-haemolytic Streptococci toxin receptor or its	
PT	fragment -	
XX		
PS	Claim 4; Page 45-47; 52pp; English.	
XX		
CC	The present sequence is full length group B beta-haemolytic Streptococci	
CC	(GBS) toxin receptor protein, HP59 from human. The present invention	
CC	relates to a method for preventing or attenuating a patho-angiogenic	
CC	condition in a mammal which comprises administering to the mammal one	
CC	or more GBS toxin receptors or their immunogenic fragments to induce	
CC	or maintain an immune response to one of GBS toxin receptors. The	
CC	method is useful for preventing or ameliorating pathoangiogenic	
CC	conditions such as cancer, scarring during wound healing, gliosis	
CC	during repair of nerve injury, chronic wounds, keloids, reperfusion	
CC	injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and	
CC	psoriasis in mammals. The proteins of the invention are also used	
CC	as vaccines.	
XX		
Sequence	536 AA:	

Dd	61	tpidpgraeaaarpccsaarynlaiibffgffiyalatrnlswalvdmvdsnttleednt	120
Qy	121	SKACEBSAPIKRVHHNQTGKKRYOMDAETOGMIGSFPEYGYITQIPGCVASKIGKMKLL	180
Dd	121	skacebsapikrvhnhqngkkywdaetogwllgsffyylltclqipgyvasklsgkml	180
Qy	181	GFGLGTAVLTLPPIAADLGVPRILVRLALEGLGECVTPPAMHAMSSNAAPLERSKLL	240
Dd	181	gfglgtavltltpriaadlgvprllvrlaleglgecvtpfamhamsswapplerskll	240
Qy	241	SISYAGAOLGTVISPLPSGIICTYMNNTYVYFEGTGTGFELMILMYLVDTPQKHKRIS	300
Dd	241	sisyagaqlgvtvisplpsgiiicymnntyvfyfgtltgfellmwlvdsdtpqkhkris	300
Qy	301	HYEKTYIILSRNQLSSQSKVPWPPIILKSUPILMAIVAHSHSYNNTFTLLTLPTYNKEI	360
Dd	301	hyeketyilsslrnqlssqskvwpwpiilksipilwaihshsynntfylltlilptymkei	360
Qy	361	LRFNNGEFGFLSSULPYLGSMICMLISGQAADNLAKKNFSTLCYRRIFSLIGMIGRAVEL	420
Dd	361	lrfnngengflsslpypgswicmlisgsqaadnlakwnfstlcyrtilfslilgmigpavel	420
Qy	421	VAAGRITGCDYSLAAVPLISTITLPGFSSGSGSIINHLDIASYAGIILGTFNEATPIGMV	480
Dd	421	vaagrltgcdyslaavflistitlpgfsssgslnhldiasyagiligtntlatpigmw	480
Qy	481	GPVIAKSLTTPNTNGEMQOTVYKIAALINVCALJFETLFARGEVONMALNDHHGRH	536
Dd	481	gpvlakslttpntngewqtvkyiaaahnvgaiftcllfarqevonwlnhghnrh	536

RESULT 3

ID	AYY45087	standard; Protein; 495 AA.
----	----------	----------------------------

AC AAY45087;

DT 31-MAY-2000 (first entry)

Partial human GBS toxin receptor (HP55).

KM Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KM pathological vasculatization; cancer metastasis; angiogenesis;
KM neovascularisation; reperfusion injury; scarring; keloid;
KM chronic inflammation; Rheumatoid arthritis; psoriasis; neural injury;
KM endothelial cell proliferation; antibacterial; anticancer;
KM anti-angiogenic; anti-inflammation; anti-arthritis; anti-psoriatic;
KM

OS Homo sapiens.

Key	Location/Qualifiers
FH	
2244	

Region	14-20
FT	/note= "Region of high hydrophilicity used to design
FT	antigenic peptide"
FT	76-90

/note= "Region of high hydrophilicity used to design

FT	Region	26..30

FT antigenic peptide" - - -

PN W0200005375-A1

PD 03-FEB-2000.

PF 22-JUL-1999; 99WO-US16676.

22-JUL-1998; 98US-0093843.

PA (UYVA-) UNIV VANDERBILT.

PI Hellerqvist CG, Fu C:

yy

DR WPI: 2000-205377/18.
DR N-PSDB: AAZ50875.
XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
XX Claim 9; Page 22; 109pp; English.
XX
CC The present sequence is partial human GBS (group B beta-haemolytic
CC streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
CC protein with seven transmembrane domains. Expression vectors comprising
CC the coding region can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
XX Sequence 495 AA:
SQ
Query Match 92.4%; Score 2621; DB 21; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.7e-261;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 42 MRSVVRDLARRDGEESTDRPILPGAPRAEAAPVCCSARNALALAFEGFTYVALRVNL 101
DB 1 mrsrvdlarrdgeestdrpilpgapraeaapvccsarnalalafgftiyalrvnl 60
OY 102 SVALVDVDSNTLEDRNTSKACPEHSAPIKVHHNOTGKKYQMDAETOGWILGSFEFGYI 161
DB 61 svalvdmvdsnttledrntskacpehsapikvhhnqtgkkyqmdaetqgwilgsffgyi 120
OY 162 ITQIPGGYVASKIGKMLGFGILGTRAVLTLPPIAADLGVPILVLRALGEGVTFP 221
DB 121 itqipggyvaskigkmlgfgilgtavltlfpriaadlgvpilvlraleglegvtfp 180
OY 222 AMHAMSSMAPPLERSKLISYAGOLGYVISPLSGIICYNNMTYVYFFETGTFW 281
DB 181 amhamsswapplersklisisyagqlgyvisplsgilcyymnmtcyvfyfgtlgtfw 240
OY 282 FILIMWLVSPTPOKHKRISHYEKEYILSLRNOLSSOKSVWPVILKSLPLMAIVAHFS 341
DB 241 filimwlvspdpokhkrihsyeykylslrnolssoksvwpvilkslplwaihahfs 300
OY 342 YNNMTFYTLTLPTMYKEILRFNVQENGFLSLPYLGSWLCMLISGOAADNLRAKNFST 401
DB 301 ynnmtfytlltllptymkeilrfnvqengflslpylgswlcmilsggaadnlraknfst 360
OY 402 LCVRRIFSLIGMIGRAVFLVAAGFIGCDYSLAAVFLTISTTLGFGSSGSINHLDIAPS 461
DB 361 lcvrrifsligmgpravlvaagfigcdyslaavfltisttlgfgssgsinhldiaps 420
OY 462 YAGILLGITWTFATIPGMVPAVLAKSLTPNTNGEMQTVYIAAIVWFGAIFFTLEPAKG 521
DB 421 yagillgitwtfatipgmvpavlaaksltpntngemqtvyyiaaivwfgaifftllepakg 480
OY 522 EVQNALNDHHGHRH 536
DB 481 evqnalndhhghrh 495
RESULT 4
ID AAB66967
XX AAB66967 standard; protein; 495 AA.
AC AAB66967;
XX

DT 18-APR-2001 (first entry)
XX
DE Human AST.
XX
KW Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
KW anion and sugar transporter; anion-cation symporter;
XX
XX Salla acid transporter.
XX
OS Homo sapiens.
XX
PN EP1069184-A1.
XX
PD 17-JAN-2001.
XX
PF 16-JUL-1999; 99EP-0202341.
XX
PR 16-JUL-1999; 99EP-0202341.
XX
PA (ALKU) AKZO NOBEL NV.
XX
DR WPI: 2001-193090/20.
DR N-PSDB: AAF55900.
XX
XX New human transporter gene implicated in Salla disease and lysosomal
PT Salla acid transport, useful in assays for identifying new drugs, or
PT diagnosing Salla acid transport defects related to mutations in the
PT transporter gene -
XX
XX Claim 1; Page 10-12; 20pp; English.
XX
XX The present sequence is human Anion and Sugar Transporter (AST) protein
CC sequence. AST has significant homology with several members of the
CC anion-cation symporter (ACS) family of transporters. AST is implicated in
CC Salla disease, and is useful in screening assays for identifying new
CC drugs. Compounds identified via AST screening is useful for preparing a
CC pharmaceutical suitable as an activator or inhibitor of a Salla acid
CC transporter protein. The pharmaceutical may be used in Salla acid
CC associated diseases and CNS/immune related disorders.
XX
XX Sequence 495 AA:
SQ
Query Match 92.4%; Score 2621; DB 22; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.7e-261;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 42 MRSVVRDLARRDGEESTDRPILPGAPRAEAAPVCCSARNALALAFEGFTYVALRVNL 101
DB 1 mrsrvdlarrdgeestdrpilpgapraeaapvccsarnalalafgftiyalrvnl 60
OY 102 SVALVDVDSNTLEDRNTSKACPEHSAPIKVHHNOTGKKYQMDAETOGWILGSFEFGYI 161
DB 61 svalvdmvdsnttledrntskacpehsapikvhhnqtgkkyqmdaetqgwilgsffgyi 120
OY 162 ITQIPGGYVASKIGKMLGFGILGTRAVLTLPPIAADLGVPILVLRALGEGVTFP 221
DB 121 itqipggyvaskigkmlgfgilgtavltlfpriaadlgvpilvlraleglegvtfp 180
OY 222 AMHAMSSMAPPLERSKLISYAGOLGYVISPLSGIICYNNMTYVYFFETGTFW 281
DB 181 amhamsswapplersklisisyagqlgyvisplsgilcyymnmtcyvfyfgtlgtfw 240
OY 282 FILIMWLVSPTPOKHKRISHYEKEYILSLRNOLSSOKSVWPVILKSLPLMAIVAHFS 341
DB 241 filimwlvspdpokhkrihsyeykylslrnolssoksvwpvilkslplwaihahfs 300
OY 342 YNNMTFYTLTLPTMYKEILRFNVQENGFLSLPYLGSWLCMLISGOAADNLRAKNFST 401
DB 301 ynnmtfytlltllptymkeilrfnvqengflslpylgswlcmilsggaadnlraknfst 360
OY 402 LCVRRIFSLIGMIGRAVFLVAAGFIGCDYSLAAVFLTISTTLGFGSSGSINHLDIAPS 461
DB 361 lcvrrifsligmgpravlvaagfigcdyslaavfltisttlgfgssgsinhldiaps 420

OY 462 YAGILGINTNTPATIPGMVPIYAKSLPDPNTVGEQWTFVYIAAINVGAIFFTLFANG 521
 DB 421 yagillgintntfatlpgmvgpviakslpdpntvgewqtfvfyiaaainvgaifftlffang 480
 OY 522 EVQNNALNDHGHRRH 536
 DB 481 evgnwalndhghrth 495
 RESULT 5
 AAM38959
 ID AAM38959 standard; Protein; 495 AA.
 AC AAM38959;
 XX
 XX 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 2104.
 XX
 XX Human; nocotropic; immunosuppressant; cytosstatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyocrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
 KW chemokineic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QH, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB: AAI58115.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 3; SEQ ID NO 2104; 10078pp; English.
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nocotropic,
 CC immunosuppressant and cytosstatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokineic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification.
 XX
 SQ Sequence 495 AA;
 Query Match 92.2%; Score 2614; DB 22; Length 495;
 Best Local Similarity 99.8%; Pred. No. 9, 2e-261;
 Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 42 MRSPPVADLARNDEESTDPTPLPGAPRAEAPVCCSARYNLAIAFEGFTVYARVNL 101
 DB 1 mrsppvrdlarndeestdptlpgapraeapvccsarynlaiafegftvayarlvl 60
 OY 102 SYALVDMVDSNTTLEDNRTSKACPEHSAPIKVHNQTKGKYOMDAETQOMIIGSFPGYI 161
 DB 61 syalvdmvdsnttlednrtskacpehsapikvhnqtkgkywdaetqvgwllsgffgyi 120
 OY 162 ITQIPGGYVASKIGCKMLGFGILGTAVLTLFTPLAADLGVPGLVLRALGEGCVTPP 221
 DB 121 itqipggyvaskigckmllgfgillgtaavtlftplaadlgvpgllvrlalleglgevtfp 180
 OY 222 AMHAMSSWAPPLERSKLLSISYAGQGTIVISLPISGILCYMMNTYFYFRFGTIGIFW 281
 DB 181 amhamsswapplerskllsisyagqgtivislpisgillcyymmntvyfyfgrtligifw 240
 OY 282 FLMTIWLVSDFPOKHKRISHYEKEYTILSLRNQSSQKSVPMVPIIKSLPLMAIVAHPS 341
 DB 241 flmtiwlvsdtpokhkrishyekeytllslrnqssqksgpwpvpiikslplmaivahps 300
 OY 342 YNMTFYTLLTLLPYWKEILRFNVQENGLSSLPYLGSM/CMILSGQAADNLRKWNFST 401
 DB 301 ynmtyflltllptywkeilrfnvqengflsslpylgswlcmllsgqaadnlrakwnfst 360
 OY 402 LCVRRTRFSILGMIGRAVFLVAAGFTGCDYSLAVAFITITTTGGPCSSSFSINHDIAFS 461
 DB 361 lcvrrtrfsligmigravflvaagftgdcyslaavafititsttggpcssfsinhdiaps 420
 OY 462 YAGILGINTNTPATIPGMVPIYAKSLPDPNTVGEQWTFVYIAAINVGAIFFTLFANG 521
 DB 421 yagillgintntfatlpgmvgpviakslpdpntvgewqtfvfyiaaainvgaifftlffang 480
 OY 522 EVQNNALNDHGHRRH 536
 DB 481 evgnwalndhghrth 495
 RESULT 6
 AAY45088
 ID AAY45088 standard; Protein; 495 AA.
 XX
 XX AAY45088;
 AC
 XX
 DT 31-MAY-2000 (first entry)
 XX
 DE Sheep GBS toxin receptor (SP55).
 XX
 KW Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55;
 KW pathological vascularisation; cancer metastases; angiogenesis; sheep;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
 XX
 OS Ovis sp.
 XX
 FH Key
 FT Domain
 FT Domain
 FT Domain
 FT Domain
 FT Domain
 FT Domain
 FT Domain
 Location/Qualifiers
 226..252
 /note= "Outer boundary of transmembrane domain"
 232..248
 /note= "Inner boundary of transmembrane domain"
 365..389
 /note= "Outer boundary of transmembrane domain"
 369..385

FT	Domain	/note= "Inner boundary of transmembrane domain"	FT	/label= PKC_phospho_site
FT	Domain	/note= "Outer boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	458..474	FT	/label= PKC_phospho_site
FT	Domain	/note= "Inner boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	135..157	FT	/label= PKC_phospho_site
FT	Domain	/note= "Outer boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	137..153	FT	/label= PKC_phospho_site
FT	Domain	/note= "Inner boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	42..58	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	328..345	FT	/label= PKC_phospho_site
FT	Domain	/note= "Outer boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	328..344	FT	/label= PKC_phospho_site
FT	Domain	/note= "Inner boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	390..407	FT	/label= PKC_phospho_site
FT	Domain	/note= "Outer boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	390..406	FT	/label= PKC_phospho_site
FT	Domain	/note= "Inner boundary of transmembrane domain"	FT	/note= "Predicted phosphorylation site"
FT	Domain	97..100	FT	/label= PKC_phospho_site
FT	Domain	/note= "Putative amidation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	59..62	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	71..74	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	77..80	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	95..98	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	225..228	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	302..305	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	357..360	FT	/label= PKC_phospho_site
FT	Domain	/note= "Asn is N-glycosylated"	FT	/note= "Predicted phosphorylation site"
FT	Domain	11..14	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	73..76	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	79..82	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	259..262	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	452..455	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	126..131	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	142..147	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	162..167	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	172..177	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	205..210	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	209..214	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	337..342	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	386..391	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	403..408	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	423..428	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	427..432	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Domain	17..19	FT	/label= PKC_phospho_site
FT	Domain	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"

FT	Modified-site	/label= PKC_phospho_site	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	37..39	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	55..57	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	73..75	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	97..99	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	254..256	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	269..271	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	276..278	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Predicted phosphorylation site"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	8..35	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Immunogenic peptide"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	7..22	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Immunogenic peptide"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	71..84	FT	/label= PKC_phospho_site
FT	Modified-site	/note= "Immunogenic peptide"	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	MO200005375-A1.	FT	/label= PKC_phospho_site
FT	Modified-site	03-FEB-2000.	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	22-JUL-1999;	FT	/label= PKC_phospho_site
FT	Modified-site	99WO-US16676.	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	22-JUL-1998;	FT	/label= PKC_phospho_site
FT	Modified-site	98US-0093843.	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	(UYVA-) UNTV VANDERBILT.	FT	/label= PKC_phospho_site
FT	Modified-site	Helleryvist CG, Fu C;	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	WPI: 2000-205377/18.	FT	/label= PKC_phospho_site
FT	Modified-site	N-PSDB; AA250876.	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	New polynucleotide encoding mammalian receptor for streptococcus toxin,	FT	/label= PKC_phospho_site
FT	Modified-site	useful for diagnosis and treatment of, e.g. pneumonia in neonates -	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	Claim 10; Page 22; 109pp; English.	FT	/label= PKC_phospho_site
FT	Modified-site	The present sequence is sheep GBS (group B beta-haemolytic	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	streptococci) toxin receptor (SP55). Sheep GBS toxin receptor is an	FT	/label= PKC_phospho_site
FT	Modified-site	integral protein with seven transmembrane domains. Expression vectors	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	comprising the coding region can be transformed into host cells to	FT	/label= PKC_phospho_site
FT	Modified-site	express GBS toxin receptor and its fragments. Detecting the receptor in	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	tissues is used to diagnose pathological vascularisation, e.g. for	FT	/label= PKC_phospho_site
FT	Modified-site	detecting cancer metastases. GBS toxin receptors are useful for treating	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	conditions associated with pathological angiogenesis or	FT	/label= PKC_phospho_site
FT	Modified-site	neovascularisation (specifically cancer, reperfusion injury, scarring	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	during wound healing, keloids, chronic inflammation (rheumatoid	FT	/label= PKC_phospho_site
FT	Modified-site	arthritis or psoriasis) or neural injury), and to raise specific	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	antibodies used for treating early onset disease. Inhibitors of this	FT	/label= PKC_phospho_site
FT	Modified-site	receptor are useful for treating pathological or hypoxia-induced	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	endothelial cell proliferation and migration.	FT	/label= PKC_phospho_site
FT	Modified-site	Sequence 495 AA:	FT	/note= "Predicted phosphorylation site"
FT	Modified-site	Query Match 82.1%; Score 2329; DB 21; Length 495;	FT	/label= PKC_phospho_site
FT	Modified-site	Best Local Similarity 86.3%; Pred. No. 2.4e-231;	FT	/note= "Predicted phosphorylation site"

Matches	427;	Conservative	39;	Mismatches	29;	Indels	0;	Gaps	0;
QY	42	MRSPVBDLARNDGEESTDTRPLLPGAPRAEAAPVCCSARYNLAIAAFGEFIYALRVNL	101						
Db	1	mkspsvdlapsdgsdgsdtrpllqrapraepapvcasaylnaiafistfgffvyslrwnl	60						
QY	102	SVALVDMVDNSTLTLEDNRTSKACPEHSAPIKYHHNOTGKKYQWDAETOGMILGSFFGYI	161						
Db	61	svalvdmvdsntlakdnrtscyaehsapikvlnhgtgkkyrwdactgwlsgffgyi	120						
QY	162	ITQIPGCIYASKIGKMLGFGILGTAVLTLPPIAADLGVGLIVLRALGEGEVTP	221						
Db	121	itqipggyvasrsgklllfiglifaatafltlfpilaadfgvalvalralegigevtyp	180						
QY	222	AMHAMSSWAPLERSKLISISYAGQIGTVISLPSGIICCYMMNTYVFFPGTIGIF	281						
Db	181	amhamsswapleriskllisisyagqigtvvslpsgyvlycymnwcyfyfifgvgllw	240						
QY	282	FLIMIVLSDTPQKHKRISHYEKEYILSLRNOLSSOKSVWVPILKSLPLMAIVVAHPS	341						
Db	241	flimivlvsdtpqkhrishyekeyilslrnolssoksvwvpilksplmaivvahps	300						
QY	342	YNNFTFTLLTLPTFYKELIREVQENGFLSLPYIGSWICMLISQOADNLRANKNFST	401						
Db	301	ywnftfylltlptfymkylrinvqengflsavpylgwlcmlisgqaadnlrarnwfst	360						
QY	402	LCVRRIFSLIGMIPVAVAGFICDYSIAVAFLTISTLTGFCSSGFSINHLIAPS	461						
Db	361	lcvrrifsligmipavfivaagfigcdyslaafltlstltgfcssgfsinhlidaps	420						
QY	462	YAGILLGINTFATIPGMVPIAKSLTPDNTVGEQWTFYIAAIVNGAIFFTLFARNG	521						
Db	421	yagillgintfatipgmipilarsltpentlgewqvfclaaainvfgaifftlfakg	480						
QY	522	EYQNMALNDHGHHRH	536						
Db	481	eyqnmalsdnhghrn	495						
RESULT 7									
ID	AAE06519	standard; Protein; 495 AA.							
AC	AAE06519;								
DT	16-OCT-2001	(first entry)							
XX									
DE		Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) protein.							
XX									
KW		Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;							
KW		cytotoxic; vulnerability; antiatherosclerotic; osteopathic; vasotropic;							
KW		prevention; attenuation; pathoangiogenic condition; cancer; scar;							
KW		wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;							
KW		keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;							
XX		vaccine.							
OS		Ovis sp.							
XX									
EH		Key							
FT		Location/Qualifiers							
FT		/label= p55a_immunogenic_peptide							
FT		/note= "Fragment of extracellular domain of							
FT		GBS toxin receptor"							
FT		9..35							
FT		/label= p56a_immunogenic_peptide							
FT		14..19							
FT		/note= "Region of high hydrophilicity"							
FT		25..30							
FT		/note= "Region of high hydrophilicity"							
FT		75..80							
FT		/note= "Region of high hydrophilicity"							
FT		71..84							
FT		/label= p57a_immunogenic_peptide							

FT		/note= "Fragment of intracellular domain of
FT		GBS toxin receptor"
XX		
PN	MO200156598-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	02-FEB-2001; 2001WO-US03662.	
XX		
PR	02-FEB-2000; 2000US-0179870.	
XX		
PI	(UYVA-) UNIV VANDERBILT.	
XX		
DR	Hellerqvist CG;	
XX		
PT	WPI: 2001-488844/53.	
PT	N-PSDB; AADI0326.	
PT	Preventing or attenuating pathoangiogenic conditions e.g. cancer,	
PT	chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by	
PT	administering group B beta-hemolytic Streptococci toxin receptor or its	
PT	fragment	
XX		
PS	Claim 8; Page 50-52; 52pp; English.	
XX		
CC	The present sequence is group B beta-haemolytic Streptococci (GBS)	
CC	toxin receptor protein, SP55 from sheep. The present invention relates	
CC	to a method for preventing or attenuating a patho-angiogenic condition	
CC	in a mammal which comprises administering to the mammal one or more GBS	
CC	toxin receptors or their immunogenic fragments to induce or maintain an	
CC	immune response to one of GBS toxin receptors. The method is useful for	
CC	preventing or ameliorating pathoangiogenic conditions such as cancer,	
CC	scarring during wound healing, gliosis during repair of nerve injury,	
CC	chronic wounds, keloids, reperfusion injury, rheumatoid arthritis,	
CC	atherosclerosis, osteoarthritis and psoriasis in mammals. The proteins	
CC	of the invention are also used as vaccines.	
XX		
SQ	Sequence 495 AA:	
Query Match 82.1%; Score 2329; DB 22; Length 495;		
Best Local Similarity 86.3%; Pred. No. 2.4e-231;		
Matches	427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;	
QY	42	MRSPVBDLARNDGEESTDTRPLLPGAPRAEAAPVCCSARYNLAIAAFGEFIYALRVNL 101
Db	1	mkspsvdlapsdgsdgsdtrpllqrapraepapvcasaylnaiafistfgffvyslrwnl 60
QY	102	SVALVDMVDNSTLTLEDNRTSKACPEHSAPIKYHHNOTGKKYQWDAETOGMILGSFFGYI 161
Db	61	svalvdmvdsntlakdnrtscyaehsapikvlnhgtgkkyrwdactgwlsgffgyi 120
QY	162	ITQIPGCIYASKIGKMLGFGILGTAVLTLPPIAADLGVGLIVLRALGEGEVTP 221
Db	121	itqipggyvasrsgklllfiglifaatafltlfpilaadfgvalvalralegigevtyp 180
QY	222	AMHAMSSWAPLERSKLISISYAGQIGTVISLPSGIICCYMMNTYVFFPGTIGIF 281
Db	181	amhamsswapleriskllisisyagqigtvvslpsgyvlycymnwcyfyfifgvgllw 240
QY	282	FLIMIVLSDTPQKHKRISHYEKEYILSLRNOLSSOKSVWVPILKSLPLMAIVVAHPS 341
Db	241	flimivlvsdtpqkhrishyekeyilslrnolssoksvwvpilksplmaivvahps 300
QY	342	YNNFTFTLLTLPTFYKELIREVQENGFLSLPYIGSWICMLISQOADNLRANKNFST 401
Db	301	ywnftfylltlptfymkylrinvqengflsavpylgwlcmlisgqaadnlrarnwfst 360
QY	402	LCVRRIFSLIGMIPVAVAGFICDYSIAVAFLTISTLTGFCSSGFSINHLIAPS 461
Db	361	lcvrrifsligmipavfivaagfigcdyslaafltlstltgfcssgfsinhlidaps 420
QY	462	YAGILLGINTFATIPGMVPIAKSLTPDNTVGEQWTFYIAAIVNGAIFFTLFARNG 521
Db	421	yagillgintfatipgmipilarsltpentlgewqvfclaaainvfgaifftlfakg 480
QY	522	EYQNMALNDHGHHRH 536
Db	481	eyqnmalsdnhghrn 495

Db	421	g l g t n f a t p n l g p l a r s l t p e n t g w t f v c a a n v f a a l f t t a k g	480
Oy	522	EVONALNDHHGHRH 536	
Db	481	evgnwalsdghghn 495	
RESULT	8		
ID	AAy45090	standard; Protein; 495 AA.	
XX	AAy45090;		
AC	AAy45090;		
XX	31-MAY-2000	(first entry)	
DT	31-MAY-2000	(first entry)	
XX	Human/Sheep	consensus GBS toxin receptor.	
DE	Human/Sheep	consensus GBS toxin receptor.	
XX	Human/Sheep	consensus GBS toxin receptor; GBS;	
KW	group B beta-haemolytic streptococci; pathological vascularisation;		
KW	cancer metastasis; angiogenesis; neovascularisation; reperfusion injury		
KW	scarring; keloid; chronic inflammation; rheumatoid arthritis; psoriasis;		
KW	neural injury; endothelial cell proliferation; antibacterial;		
KW	anticancer; anti-angiogenic; anti-inflammatory; anti-arthritis;		
KW	anti-psoriatic.		
XX	Homo sapiens.		
OS	Ovis sp.		
XX	Key	Location/Qualifiers	
FT	Protein	1..495	
FT		/note= "All the X's in the sequence correspond to	
FT		non-consensus sites"	
XX	WO200005375-A1.		
PN	03-FEB-2000.		
XX	22-JUL-1999;	99WO-US16676.	
PD	22-JUL-1999;	99WO-US16676.	
XX	22-JUL-1998;	98US-0093843.	
PF	(UYVA-) UNIV VANDERBILT.		
XX	Heilerqvist CG, Fu C;		
PI	WPI: 2000-205377/18.		
XX	N-PSDB: AAZ50880, AAZ50881.		
DR	New polynucleotide encoding mammalian receptor for streptococcus toxin,		
PT	useful for diagnosis and treatment of, e.g. pneumonia in neonates -		
XX	Disclosure; Page 98-100; 109pp; English.		
PS	The present sequence is the human/sheep consensus GBS (group B beta		
CC	-haemolytic streptococci) toxin receptor. Expression vectors comprising		
CC	the coding region can be transformed into host cells to express GBS		
CC	toxin receptor and its fragments. Detecting the receptor in tissues is		
CC	used to diagnose pathological vascularisation, e.g. for detecting cancer		
CC	metastases. GBS toxin receptors are useful for treating conditions		
CC	associated with pathological angiogenesis or neovascularisation		
CC	(specifically cancer, reperfusion injury, scarring during wound healing,		
CC	keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or		
CC	neural injury), and to raise specific antibodies used for treating early		
CC	onset disease. Inhibitors of this receptor are useful for treating		
CC	pathological or hypoxia-induced endothelial cell proliferation and		
CC	migration.		
XX	Sequence	495 AA;	
XX			
XX			

Query Match	59.9%;	Score 1698;	DB 21;	Length 495;
Best Local Similarity	67.8%;	Pred. No. 3.3e-166;		

[illegible]

RESULT	9
AA65238	
ID	AA65238 standard; Protein; 314 AA.
XX	
AC	AA65238;
XX	
DE	20-NOV-2001 (first entry)
XX	
DE	Human sodium dependent phosphate co-transfer protein 35.
XX	
KM	Human; sodium dependent phosphate co-transfer protein 35;
KW	hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW	nephritis; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	CNI298882-A.
XX	
PD	13-JUN-2001.
XX	
PF	06-DEC-1999; 99CN-0124217.
XX	
PR	06-DEC-1999; 99CN-0124217.
XX	
PA	(SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX	
PI	Mao Y, Xie Y;
XX	
DR	WPI; 2001-503367/56.
DR	N-PSDB; AAH79234.
XX	
PT	Human Na-dependent phosphate cotransporter 35 and its coding sequence -
PS	Claim 1; Page 21(Disclosure); 28pp; Chinese.

XX The present invention provides the protein and coding sequences of human
 CC sodium dependent phosphate co-transfer protein 35. The sequences can be
 CC used in the treatment of hypophosphaturia, hypercalcaemia,
 CC hypophosphataemic rickets and nephritis. The present sequence is the
 CC protein of the invention.

XX Sequence 314 AA:

Query Match 59.2%; Score 1679; DB 22; Length 314;
 Best Local Similarity 100.0%; Pred. No. 1,6e-164;
 Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 MHAMNSMAPPLERSKLTISYAGAGLTIVISLPLSGIICYYNMWYVFEGTIGIFWE 282
 DB 1 mhamsmapplerskltisysagagltivislpisgllicymnmwtyvfegtlgltf 60
 QY 283 LMIWLVSDTPQKRRISHYEKEYILSLRNOLSSQSVWPVPIKSLPLMAIVAHFSY 342
 DB 61 lmiwlvsdtpqkrrishyekeyilslrnqlsqsvwpvpilkslplmavahfsy 120
 QY 343 NMTFTYLLPLPYKMEIRFNVQENGFTSSLPYIGSNLCMILSGAALNLRKMFSTL 402
 DB 121 nmtftyllplpykmeirfnvqengftsslpyslcmilsgaalnlrkwmfstl 180
 QY 403 CVRRISLIGMIGPAFLVAAGFIGDYSIAVAFLITFTLGCSSGFSINHLDIAPSY 462
 DB 181 cvrrisligmigpavflvaagfigcdyslavaflltftlgcsgfsinhldiap 240
 QY 463 AGILGINTFATIPGMVGVIAKSLTPDNTVGEWQTVFYIAAINVFGAIFETFAKE 522
 DB 241 agilgintfatipgmvgvialksltpdntvgewqtvfyiaainvfgaifetfakge 300
 QY 523 YONWALNDHGHNRH 536
 DB 301 yonwalandhghnrh 314

RESULT 10

AAM40745
 ID AAM40745 standard; Protein; 309 AA.

XX AAM40745;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 5676.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0532317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

XX Tang YR, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR N-PSDB; AA159901.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 5676; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S. disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 309 AA:

Query Match 56.9%; Score 1614; DB 22; Length 309;
 Best Local Similarity 100.0%; Pred. No. 7.8e-136;
 Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PPPVOFAPRPGFGISGRSLICQVASTPAHVGVMSPVYDLARNDESTDRPLPGA 67
 DB 6 pppvofaprpfgfgisgrslleqvastpahvwmsspyrdlarndeestdrpllpga 65
 QY 68 PRAEAPVCCSARKYNLAIAFFGFTVYALRVNLGYALVDMVDSNTTLEBNRTSKACPEH 127
 DB 66 praeapvccsaraynlaiiaffgftvayalrvnlsvayldvmdsnttlebnrtskacpeh 125
 QY 128 SAPIKVHNQGTGKYWDMEETQGMILGSFFYGYITQIDREGVYASKTGKMLIGFSLGT 187
 DB 126 sapikvhnqgtkkywdmetqgmilgsffgyitqidregvyasktgmllgfglgt 185
 QY 188 AVLTLETPRAADLGVGRLLVLRALBGLGCVFRRANHAMSSWAPPLERSKLTISYAGA 247
 DB 186 avltletpraadlgvgrllvrlalbgldgcvfrranhamsswapplerskltisysaga 245
 QY 248 QLGTVISLPLSGIICYYNMWYVFEGTIGIFWFLMIWLVSDTPQKRRISHYEKEYI 307
 DB 246 qlgtvislplsgllicymnmwtyvfegtlgltfllwlvsvdtpqkrrishyekeyi 305
 QY 308 LSSL 311
 DB 306 lssl 309

RESULT 11

AAM93914
 ID AAM93914 standard; Protein; 277 AA.

XX AAM93914;

XX 06-NOV-2001 (first entry)

XX Human polypeptide, SEQ ID NO: 4069.

XX

Db	324	egqdevsirapbeeipwsslltvsyplvailltqcqgwafyrlqtlelptysnnllhfdi	383
Oy		QENGFLSSLPYLGSMLCMILSGQADNLRKNNFSTLCRRRFSILGMIGPAVFLVAAGF	425
Db	384	gnallnaavpylltswfvgiaccalsadlmdrlaryisilnsyxlwntvasvvpelgllgily	443
Oy	426	IGCDLSLVAFLFTISTLGGFCSSGSFSTNHLDIAPSYAAILLGITRFTFTFGMGP-VI	484
Db	444	vgcdvrvvtfmlagvsgfigavaygnqmhlaIspryagtlmyglttsaanlcgflagpyl	503
Oy	485	AKSLPDMNTGEMGTVEFYIAAIIINVGALFPLLPFAKGEYONNA	527
Db	504	gllinhretlqwhlvrtwlaaglnlaagnfiylilfaaeqsws	546
RESULT	13		
ID	ABB60525		
AC	ABB60525	standard; Protein; 502 AA.	
XX	ABB60525;		
XX	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster	polypeptide SEQ ID NO 8367.	
XX			
XX	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US09231.		
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX			
PA	(PERK) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX	WPI; 2001-656860/75.		
DR	N-PSDB; ABL04628.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		
XX			
PS	Disclosure; SEQ ID NO 8367; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (AB101840-AB130511), expressed DNA		
CC	sequences (AB17737-ABB72072).		
CC	(ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
XX	Sequence 502 AA;		

Query Match	35.0%	Score 992;	DB 22;	Length 502;
Best Local Similarity	42.4%;	Pred. No. 2.5e-93;		
Matches 190; Conservative	93;	Mismatches 149;	Indels 16;	Gaps 3
OY	80 RYNAILAFEGGFIVYALRNVISVALVDVMDNSTLTLEDNRTSKACDEHSAPTKIVHHNQTG	139		

```

Db      38  tyivvllaiffgfifvnylsrvnlsvaivaamteurtvfd-----adgnvysyq-- 83
Oy      140  KKYOWDAETOCMIWGSEFFYGIITQIPEGYVASKIGGKMLGFILGTAVLTFTTPIAAD 199
Db      84  -dfpdsxvqklllssffiyyltqlfgyygtklygnvltvgctagctalltltcmaas 142
Oy      200  LGVGLIYLRALEGGEGEVTPRPAHNAWMSWAPPLERSKILSISYAAQAGTVYISPLSG 259
Db      143  hslmfivrlvrliegffegvttprghavawatspplersmasla fagynyagvvaampcs 202
Oy      260  IICVYMMNTYVFEFGTIGIFWFLMIWLWISDPQOKHRISHYKEYILSLRNOLSSQK 319
Db      203  flactygnestfyfytgtlqvlywtlvtvfkagpelirfcskeecdyiqctlygyvskhv 262
Oy      320  SVPWVPIILKSLPLMAIVAHPSYMMTFYTLTLPLTYMKELIRENNOENFSLSLPILGS 379
Db      263  khpwaiftcmprfyaashashsewmgfylltqtprsftrdlnnddykglbsavpylam 322
Oy      380  WCMWILSGOANDNLRAKNWSESLCVRRIEFLIGMIGAPVFLVAAGTGCDYSLAAVFLTT 439
Db      323  gillavsygladwlvqkvkwtlctqrnnfocagelaqvtlmmlltayl-lqptsvsallt 381
Oy      440  STTLGGFCCSSGFSINHDIASVYAGILGINTNTATIPGMVPIAKSLPNDTVGEMQT 499
Db      382  avgjgafawsgfayvnhldiapqhasvimgjntfatipjvslplltgylvntqgsdewrl 441
Oy      500  VFYIAAIAINVGAIFFTLFAKGEOONNA 527
Db      442  iftsagilyvgcvlywfyccgdjwena 469

```

RESULT 14
ABB65873 standard; Protein; 502 AA.
ID ABB65873 standard; Protein; 502 AA.
XX
AC ABB65873;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 24411.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX OS
XX WO200171042-A2.
XX FM
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-061415O.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX WPI: 2001-656860/75.
XX DR N-PSDB: ABLO9976.
XX DT
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 24411; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention

Search completed: July 15, 2002, 08:24:01
Job time: 106 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 09:39:25 ; Search time 4323.72 Seconds
(without alignments)
9146.309 Million cell updates/sec

Title: US-09-776-865-1
Perfect score: 2930
Sequence: 1 gtctgcgtcgaagccctccccc.....ctcaaaaaaaaaaaaaa 2930

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13376207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	972	33.2	985	9	AL550137
2	775.4	26.5	787	9	BG198416
3	712.6	24.3	754	10	BI907284
4	710	24.2	729	10	BF971208
5	648.2	22.1	700	10	BE869819
6	626.2	21.4	689	10	BG400588
7	610.2	20.8	665	10	BG290613
8	609.6	20.8	676	10	BE867611
9	599.6	20.5	626	10	BF02084380
10	575.6	19.6	661	10	BG400668
11	572.2	19.5	736	10	BI860521
12	554.6	18.9	931	10	BE878141
13	524	17.9	841	10	BF312727
14	523.8	17.9	801	10	BI661062
15	521	17.8	524	10	BM272092
16	491.2	16.8	499	9	AL597124
17	477.6	16.3	711	10	BI697765

Result No.	Score	Query Match	Length DB	ID	Description
18	462	15.8	462	9	AI075876
19	445.2	15.2	904	10	BF124137
20	438.2	15.0	895	10	BF539146
21	432	14.7	432	9	AA258513
22	423.8	14.5	427	10	BM271768
23	420.2	14.3	682	9	BB613552
24	413	13.9	579	10	N21319
25	406.8	13.7	411	9	AA258357
26	402.8	13.5	633	10	BI817031
27	394.8	13.1	412	9	AM293672
28	385.2	13.1	645	9	BB537525
29	384.4	13.1	458	10	H71642
30	383.8	13.1	518	9	AA833297
31	381.8	13.0	630	10	BG541099
32	381	13.0	381	10	BF772376
33	367.8	12.6	372	9	AI081147
34	367.2	12.5	629	9	BB664731
35	353.4	12.1	361	9	AI337429
36	349.4	12.1	893	10	BG541099
37	349.4	11.9	351	9	AI245423
38	348	11.9	349	10	BF589677
39	345.8	11.8	450	10	H71643
40	342.4	11.7	1100	10	BG866603
41	318	10.9	329	9	AI079154
42	316.4	10.8	429	10	BF563945
43	307.8	10.5	367	10	H63685
44	301	10.3	301	10	N31254
45	288.6	9.8	721	10	B060500

ALIGNMENTS

RESULT 1
AL550137 1
LOCUS AL550137 LTI_NFL006.PL2 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1040XR17 5
PRIMER, mRNA sequence.
ACCESSION AL550137
VERSION AL550137.1 GI:12886813
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers
1..985

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1040XR17"
/clone_lib="LTI_NFL006.PL2"

/note="Vector: pcwvSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcwvSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com url :
http://fulllength.invitrogen.com"

BASE COUNT 236 a 224 c 224 g 300 t 1 others
ORIGIN

Query Match 33.2%; Score 972; DB 9; Length 985;
 Best Local Similarity 99.7%; Pred. No. 3.2e-187;
 Matches 983; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 327 tftcggagcgccgctccctctctctgccaagtgccagtaacctgtcgaatgagcgta 386
 |||||||
 Db 1 tftcggagcgccgctccctctctctgccaagtgccagtaacctgtcgaatgagcgta 60

QY 387 tgaagctccggtctcgaacctgcccgcgaagatgcgcgagagagacagcgacgc 446
 |||||||
 Db 61 tgaagctccggtctcgaacctgcccgcgaagatgcgcgagagagacagcgacgc 120

QY 447 ctctctaccgagcgccccaagcgccgaagcgccgcagtgctgctctgctcttaca 506
 |||||||
 Db 121 ctctctaccgagcgccccaagcgccgaagcgccgcagtgctgctctgctcttaca 180

QY 507 acttaagaatttggccttttgggtttctctctctctctctctctctctctctctga 566
 |||||||
 Db 181 acttaagaatttggccttttgggtttctctctctctctctctctctctctctctga 240

QY 567 gttgctgattgagatgtagatcaatacaacttagaagataagaactcca 626
 |||||||
 Db 241 gttgctgattgagatgtagatcaatacaacttagaagataagaactcca 300

QY 627 aggcgtgtccagagcaattctctcccaataaagttcataaaacgagtgtaagat 686
 |||||||
 Db 301 aggcgtgtccagagcaattctctcccaataaagttcataaaacgagtgtaagat 360

QY 687 accaattggagatgagaagaactcaagaatgattctggttccctttttatgctaca 746
 |||||||
 Db 361 accaattggagatgagaagaactcaagaatgattctggttccctttttatgctaca 420

QY 747 tccacaagattctctgaggaatagttgtccagcaaaatagaggggaaatgctgtagat 806
 |||||||
 Db 421 tccacaagattctctgaggaatagttgtgtccagcaaaatagaggggaaatgctgtagat 479

QY 807 ttggagatccttggcaactgtctccaccctgttcaactcccatgtgtcgaattagag 866
 |||||||
 Db 480 ttggagatccttggcaactgtctccaccctgttcaactcccatgtgtcgaattagag 539

QY 867 ttggagcaactcattgtactcagaagcaactgaagaagtagagaggtgttactccag 926
 |||||||
 Db 540 ttggagcaactcattgtactcagaagcaactgaagaagtagagaggtgttactccag 599

QY 927 ccatgcatgcatgtgtctctcttgggtcccccctcttgaagaagaacactcttaga 986
 |||||||
 Db 600 ccatgcatgcatgtgtctctcttgggtcccccctcttgaagaagaacactcttaga 659

QY 987 ttctgatatgagagacagcttggagacagtaattctctctctctctctctctctga 1046
 |||||||
 Db 660 ttctgatatgagagacagcttggagacagtaattctctctctctctctctctctga 719

QY 1047 gctacatatgaattggtacttaagtctctctctctctctctctctctctctctga 1106
 |||||||
 Db 720 gctacatatgaattggtacttaagtctctctctctctctctctctctctctctga 779

QY 1107 tctcttctgagatcgttagtagtagacacaccacacacacacacacacacacacac 1166
 |||||||
 Db 780 tctcttctgagatcgttagtagtagacacacaccacacacacacacacacacacac 839

QY 1167 atgaaagaagaatacattctctcatcatlaagaatcagcttctctcagaagtcagtg 1226
 |||||||
 Db 840 atgaaagaagaatacattctctcatcatlaagaatcagcttctctcagaagtcagtg 899

QY 1227 cgtgggtaccattttaaaatcccgccacttgggcatcgttagtgcacatttctt 1286
 |||||||
 Db 900 cgtgggtaccattttaaaatcccgccacttgggcatcgttagtgcacatttctt 959

QY 1287 acaactggaactttatattattatg 1312
 |||||||
 Db 960 acaactggaactttatattattatg 985

RESULT 2
 BG198416
 LOCUS R5717674 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG198416
 ACCESSION BG198416
 VERSION BG198416.1 GI:13720103
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 787)
 Harrington,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
 Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J.,
 Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,
 E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,
 J., Danzig,J., and Ducar,M.
 Creation of genome-wide protein expression libraries using random
 activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 JOURNAL 21227151
 MEDLINE
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave., Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scaine@atersys.com
 High quality sequence stop: 549.
 Location/Qualifiers
 1..787
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 214 a 166 c 154 g 253 t
 ORIGIN

Query Match. 26.5%; Score 775.4; DB 10; Length 787;
 Best Local Similarity 99.2%; Pred. No. 2.8e-147;
 Matches 779; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1645 gtagctgtatcctctctgagcaacaatacattgccaattccaggaatgttg 1704
 |||||||
 Db 3 gtagctgtatcctctctgagcaacaatacattgccaattccaggaatgttg 62

QY 1705 gccgcgtcattgctaaagtcctgaaccctgaatacaactgttggagaatggcaaccgtgt 1764
 |||||||
 Db 63 gccgcgtcattgctaaagtcctgaaccctgaatacaactgttggagaatggcaaccgtgt 122

QY 1765 ccatattgtctgtctatgaatttgggtccattttcttaactatccgaagg 1824
 |||||||
 Db 123 ccatattgtctgtctatgaatttgggtccattttcttaactatccgaagg 182

QY 1825 tgaagtacaaacttggctcctcatgatacattgacagacagacactgaagaaacata 1884
 |||||||
 Db 183 tgaagtacaaacttggctcctcatgatacattgacagacagacactgaagaaacata 242

QY 1885 aataacacctgcctcatlaatgatttattatcatgtaaacctcaagtcctctgt 1944
 |||||||
 Db 243 aataacacctgcctcatlaatgatttattatcatgtaaacctcaagtcctctgt 302

QY 1945 attgtgaagcatttaagcttttttaattgtaactgtatgaatttaagccat 2004
 |||||||
 Db 303 attgtgaagcatttaagcttttttaattgtaactgtatgaatttaagccat 362

Oy	2005	aacttggaataatcacctaggtttgcgaaataataataaactcgtttaaattatbaataa	2064
Db	363	AATCATGAAATPACCTAGTTCGAGAATAAATAAACCTGTCTTAATTATGAATPA	422
Oy	2065	tatgaacctaaggaccttacttacttagttcaacatacccgccctgctagtccggacaatga	2124
Db	423	TATGTAACCTTAGGACTTTACTTTAGTTCACATACCTGCCTGCTAGTCGGCAACATGA	482
Oy	2125	agtagaacagtcctctgttatctttttaagggcacatacctaaaggaatgaagctgaacaagacc	2184
Db	483	AGTAGAGACAGTTCCTGTTGATTTTTAGGGCCATACTAAAGGAATGAGCTGAAACAGACC	542
Oy	2185	tccgtatacccttctgtctaattaactagaatgatatttcaggtactgataaacaacctgt	2244
Db	543	TCTCGATCACCTTTGGTTTAATTAACATAATATATATCTCAGGTACTATTAACACCTGT	602
Oy	2245	tgttgttaaccttccctaataaaaattgfcagcctcctcctctgcaacttagaacccaacttt	2304
Db	603	TGTTGTTACCTTTTGCTCATAAAAATTGTCAGCTCTCTGTGACGCTTAACCTCAAACTTT	662
Oy	2305	agcatctctgfyagagctgycacatccactgylataahttcgcccctgycacaactgyactga99ga	2364
Db	663	AGCATCTCTGTGAGAGCTGCCATCCACTGTATAATTTTCCTGCGCAACTGAGACTGAGGGGA	722
Oy	2365	gtfggccagagcagcagctgcgaagcaactccccctccctggtctcaaggtgcagaatgcccagcgt	2424
Db	723	GTTGCCAGAGCAGCTGCACAAGCACTCCCTTCCTGGCTTCAGGGGTCAAGAGTGCACAGTGT	782
Oy	2425	ttatc 2429	
Db	783	TTATC 787	
RESULT	3		
LOCUS	B1907284	754 bp	mRNA linear EST_16-OCT-2001
DEFINITION	60306385FP1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:521276 5'		
ACCESSION	B1907284		
VERSION	B1907284.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mhc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabds@email.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLU) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMLU at: http://image.llnl.gov Plate: LLM11534 row: d column: 15 High quality sequence stop: 750. Location/Qualifiers 1. 754 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5212766" /clone_lib="NIH_MGC_118" /tissue_type="Leukocyte" /lab_host="DH10B" /note="Vector: pCMV-Sport6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA source: leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range		
FEATURES	source		

1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH/MGC Library."

Query Match	24.3%	Score 712.6	DB 10	Length 754
Best Local Similarity	99.1%	Pident. No. 1.6e-134		
Matches 748	Conservative 0	Mismatches 4	Indels 3	Gaps 3
QY 310	gccccggggcttcggcgtctcggccggccgctccctctctcgcagctggcgagctacac	369		
Db 2	GGCCGGGGGCTTCGGGCTGTGGGGCGGGCGCTCCTTCTTGCCAGGTGGGAGTACAC	61		
QY 370	tgtctcagtagcgtlcaltagagctccggttcagacctggtcccgaaagatgycgagga	429		
Db 62	TGCTACACTAGCGCTCAGAGAGGTCTCCGGTTCGAGACCTGGCCCGGAACGATGGCGAGGA	121		
QY 430	gagacaggaacgcagccgctctctacaggggcccccaaggccgaagccgctccagtg	489		
Db 122	GAGCAGGACCGC ACGGCTCTTTTACGGGGGGCCCCACGGGGCCGAGCCGCTCACGTTC	180		
QY 490	ctgtctctgctcgtttacaaacttagcaaatcttgaccttttttggtttcttcatgtglatgc	549		
Db 181	CTGCTCTGCTCGTTACCAACTTAGCAATTTTGGCCCTTTTGGTTCTTCATTTGTGTATGC	240		
QY 550	atttcgtgtgaatctgtagtgttgctgtagtggatagtgatgttagatccaacttaag	609		
Db 241	ATTACGTGTGAATCTGAATGTTGGCTTATGTGGATATGTGTAATTCMAATACAACTTTGA	300		
QY 610	agataatgaactctccaaagcgctgcagagcatctctccataaaagtcatcacaa	669		
Db 301	AGATATATGAACCTTCCAGGCGGTGTCAGAGCATTTCTCTCCATATAAATTTCATCATRA	360		
QY 670	tcaaacgggtaagaagtaaccaatggagatgcagaaactcaagatgattctcgtctct	729		
Db 361	TCAACGGGTAAAGATGCCAATGGGATGACAGAACTCAAGATGATTTCTGGTCTCT	420		
QY 730	ttttatagctacatcatcacacagattctctggagagatagtctgcagcaaaatagggg	789		
Db 421	TTTTATATGGCTACATCATACACACAGATTCTTGAGAGATATGTTGCCACCAAAATAGGGGG	480		
QY 790	gaaatgctgttagatttgggaatcctctggacgtgttctctaacccgttcaaccctat	849		
Db 481	GAAATGCTGTAGCGCATTTGGATCTCTGGACATGCTCTCTCCACCGCTTCTACCTCAT	540		
QY 850	tgctcgagatcttagaggttggaccactcatctgtaactcagaaactaagaagactagaga	909		
Db 541	TGCTGACAGATTTAGGAGTGGACCACTGATTTGTACTAGACACATRAAAGGACTAGAGA	600		
QY 910	gggtgttatacatctcagcgaatgatgcagtgtgtgtctctctgggtcccccctctgaag	969		
Db 601	GGGGTTTACATTTCCAGGCATGACATGCAATGTGGCTTCTGGGGCTCCCTCTTGAAG	660		
QY 970	aagcaaaccttcttagcatctgtatgcaaggagcaagcttggaaagtaaa-tttctctc	1028		
Db 661	AAGCAAACTTTTACATTTTATATGACAGAGACACAGCTTGGGACAGATATTTTCTCTTC	720		
QY 1029	ctctctctgataaattgctactataatgaaattgg 1063			
Db 721	CTCTTTTGGAAATAA-TTGTACTATATGAAATTGG 754			

RESULT 4

BF971208

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

BF971208

6022273220F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4361140 5', mRNA

BF971208

BF971208.1 GI:12338423

729 bp

linear

EST 22-JAN-2001

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: LLM10003 row: h column: 05
High quality sequence stop: 699.
Location/Qualifiers
1. 729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4361140"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally; Oligo df: primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."
BASE COUNT 206 a 149 c 136 g 238 t
ORIGIN
Query Match 24.2%; Score 710; DB 10; Length 729;
Best Local Similarity 100.0%; Pred. No. 5.5e-134;
Matches 710; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1669 cacaataacttgcacattccagaagatggttgcccgctattgtctaaagctgcac 1728
DB 1 CACAAATACATTTGCCATTCAGGAATGTTGGGCCGCTATTCCTAAAGGTCTGAC 60
QY 1729 cccgtataaacactgttgagaatgcgcaaccgctctatctatgcctgctataatgt 1788
DB 61 CCCGATTAACACGTGTGGAATGCAACCGTCTATATGCTGCTCTATTAATGT 120
QY 1789 tttgtgtgccatttcttctacatttcgcaaaagtgaaagtaacaaactgggctctcaa 1848
DB 121 TTTTGGTCATTTCTTTACATATTCGCCAAAGGTACAAACAACTGGGCTCTCAA 180
QY 1849 tgaaccacatggacacagacactggaagaaacaaataatctctgcctctattaatgta 1908
DB 181 TGATCACCATGACACAGACACTGAGAGAACCAATAATATATCTGCTCTATTAAAGTA 240
QY 1909 ttcttatctatgtaaacctcaaaagtgccttctgtatctgtgaagatcatatgtcttt 1968
DB 241 TTTTATTTATTCATGTAACTCAAGTCCCTTGATATGTGTAAGATCTATATGCTTT 300
QY 1969 tttaattgacttgatagatttttaagacctataatcaatgaataactagttgcc 2028
DB 301 TTTTATTTGACTTGTATGTATGATTTTAAAGCCTATATATCATGAATAATCATAGTTGCC 360
QY 2029 agataataaataagactggttaataatgaataataatgaagctagagctctacttt 2088
DB 361 AGATATATAAATAAGAACTGTTTAATATGAATAATATGTAAAGCTAGAGCTTCTACTTT 420
QY 2089 aggtcacatacctgcctgctagtcggaacatagaatgaagacagcttctgtgatttt 2148
DB 421 AGGTTCACATACCTGCTGCTAGTCGGGCAACATGAAGTAGACAGTGTCTGTGATTTT 480
QY 2149 tagggccataactaaaggaatgagctgaaacagacctcctgataccttgccttaataa 2208

DB 481 TAGGCGCATCTAAAGGAGATGAGCTGAACACACCTCTGATACCTTTGCTTAATTAA 540
QY 2209 ctgataataattctcaaggtacacgataaacaactgtgtgttctacttctctataaaaa 2268
DB 541 CTGATGATATATTTCTCAGGTACGTGATTAACACCTGTTGTTGTTCTTCTATATAA 600
QY 2269 ttgtcagctctctctgacacttagacctcaacttagacatctctgtgagctgcaccc 2328
DB 601 TTGTCACCTCTCTGTGACATTTAGACTCAACTTTATAGCATCTCTGTGAGAGTCCATCC 660
QY 2329 actgtataattgcctgacactggaactggaaggaggtgtgccagcag 2378
DB 661 ACTGTATAATTTGCGCTGCGCACTGAGAGGAGGTGTGCCAGGACAG 710
RESULT 5
BE869819 700 bp mRNA linear EST 20-OCT-2000
LOCUS 60144651F1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3850740 5',
DEFINITION mRNA sequence.
ACCESSION BE869819 GI:10318595
VERSION BE869819.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: LLM9570 row: e column: 13
High quality sequence stop: 697.
Location/Qualifiers
1. 700
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3850740"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo df. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT 185 a 141 c 130 g 244 t
ORIGIN
Query Match 22.1%; Score 648.2; DB 10; Length 700;
Best Local Similarity 99.1%; Pred. No. 2e-121;
Matches 683; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 1408 tatgaccctgctgctgcacagctgctacatttaagggcaaaatggaattttcaactt 1467
DB 1 TATGATCTCTGCTGTGCAAGCTGCTGACAAATTTAAAGGCAAAATTTTCAACTTT 60
QY 1468 atgtgtcgcaagaattttagccttataagaatgactggaactgcaglatctcgtgtagc 1527
DB 61 ATGTGTCCGAGAAATTTTAAAGCTTATAGGAATGATTTGACCTGCAGTATTCCTGTATAC 120
QY 1528 tgctgctcatttgctgtgattatctcttggcggctgtcttctcaatataatcaaacac 1587

Db 121 TCGTGGCTTCATGGCTGTGATTAATTCCTTTGGCCGCTGCTTCTTAACATATACAAAC 180
QY 1588 actgggaagccttct 1647
Db 181 ACTGGAGGCTTTGCT 240
QY 1648 tgcctgtaaccctccgagcgcac 1707
Db 241 TCGTGTATCCCTCCGCGGATCACAATATCATTTGCCACTATTCAGAGATGTTGGGCC 300
QY 1708 cgtcatctgtaaaagctcgaacccctgataaacctgttggagaatgycacaacccgttctta 1767
Db 301 CCGTATGCTAAAGCTGACCCCTGATTAACACTGTGTGAGATGGCAAAACCGTGTCTTA 360
QY 1768 tatctct-gctgctcttaagtcttctctctctctctctctctctctctctctctctctct 1826
Db 361 TATTCTCTCCTCTCTATTAATGTTTGGGCGCATTTCTTAACCTATTCGCGCAAGGGT 420
QY 1827 aagtaaaaactcggcctcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaat 1886
Db 421 AAGTCAAAAGCTGGCTCTCATGTATGACCAATGACAGACAGACAGACAGACAGACAGACATAA 480
QY 1887 taatccctcct 1946
Db 481 TAATCCTGCTCTATTAATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 540
QY 1947 tctgaagcct 2006
Db 541 TGTGTAAGCATCTATGCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 599
QY 2007 tcatgaatatcact-actgtccagaataataataataataataataataataataataataata 2065
Db 600 TCATGAATATATACACAGATGTCGACAGATATTAATTAATGAACCTGTCTTCTTATGATATAT 659
QY 2066 atgtaagctaggaactctcaacttaagttc 2094
Db 660 AAGTAAGCTAGGACTTCTTAACCTTAGGTTTC 688

RESULT 6
LOCUS BG400588 689 bp mRNA linear EST 12-MAR-2001
DEFINITION 602464241P1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592171 5',
mRNA sequence.
ACCESSION BG400588
VERSION BG400588.1 GI:13294036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 689)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rsraus@nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LINC1330 row: b column: 12
High quality sequence stop: 687.
Location/Qualifiers
1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592171"
/clone_id="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"

FEATURES

source

/note="Organ: kidney; Vector: PDNR-LIB (Clontech); Site 1:
SfiI (ggcggccggcc); Site 2: SfiI (ggcctatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCATTAATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGAGCGCGCGCATG-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 193 a 134 c 135 g 227 t
ORIGIN
Query Match 21.4%; Score 626.2; DB 10; Length 689;
Best Local Similarity 98.1%; Pred. No. 5.8e-117;
Matches 676; Conservative 0; Mismatches 8; Indels 5; Gaps 4;
QY 1688 attcagaatgtgttggccgctcattgtctaaagctcgaacccctgataacactgttga 1747
Db 1 ATTCAGAGATGTGTGGGCCCTCATGTCTAATAAGTCACCCCGATTAACACTGTGGA 60
QY 1748 gaatggcaaacccgtgtctctatctgtctctctctctctctctctctctctctctctct 1807
Db 61 GAATGGCAAAACCGTGTCTATATGCTCTGCTATTAATGTTTGGGCCATTTCTTT 120
QY 1808 acactattcgcgaagtggaagtaacaaactgtctcactgaatcaccatgacacga 1867
Db 121 ACATATTCGCGCAAGGTGAAGTACAAAACCTGGCTCTCAATGATCACCATGACACACA 180
QY 1868 cactgaagaaacaaataataatcctgtccctctctctctctctctctctctctctctctct 1927
Db 181 CACTGAAGAACCAATTAATAATCTGCTCTATTAATGATTTTATTAATCAATGAAC 240
QY 1928 ctcaaatgtcctctgtatctgtgaagcattctatgtccttctctctctctctctctct 1987
Db 241 CTCAAGTGCCTTCTGTATGTTGTAAGCATCTATGTTTATTAATGATGTTGAT 300
QY 1988 agattttaagcctataatacagaataataataataataataataataataataataata 2047
Db 301 AATTTTTAAAGGCTATATATCATGAATATACATGTTGCGCAATTAATAATGAATAC 360
QY 2048 tcttaatatataataatgaagctgaagctgaagctgaagctgaagctgaagctgaagct 2106
Db 361 TGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
QY 2107 gctagtcgggcaacatgaagtagaacagttctgt-tgatttttagggccataactaaag 2165
Db 421 GCTAGTCGGGCAACATGAAGTAGGACAGTCTGTGATTTTAAAGGCGCATTAAGG 480
QY 2166 gaatgaagctgaagcactcctgtctctctctctctctctctctctctctctctctctct 2225
Db 481 GAATGAGCTGAAGACCTCTGTATACCTTGTGTTAATTAAGATGAATTAATTCGA 540
QY 2226 ggtactgaataaacctgtgtgtctcaact--tctcctaataaatgttcaactcct 2283
Db 541 GGTACTGATTAACACCTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600
QY 2284 gacacttagacccaacttta-gaatctctgttgaagctgtccatccactgtataatctg 2342
Db 601 GACACTTAGACCTCAACCTTAGGATCTGTGTGAGAGCTGCATCCACAGTATTAATTCG 660
QY 2343 cctgtgaactgtgaagtgagtggtgc 2371
Db 661 CCGTGAACCTGAGCTAGAGGAGAGTGTGCC 689

RESULT 7

LOCUS BG290613 965 bp mRNA linear EST 21-FEB-2001
DEFINITION 602388889P1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4517662 5',
mRNA sequence.
ACCESSION BG290613

VERSION	EST.	1	GI:13047751
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 965)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNT) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNT at: http://image.llnl.gov Plate: LLAM10411 row: a column: 23 High quality sequence stop: 634. Location/Qualifiers		
FEATURES	1..965		
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4517662" /clone_id="NIH_MGC_93" /tissue_type="transitional cell papilloma, cell line" /lab_host="DH10B (phage-resistant)" /note="organ: bladder; Vector: PCMV-SPOrt6; Site:1; NCBI; Site:2: Salt; Cloned unidirectionally; oligo-dr primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	281 a 218 c 223 g 243 t		
ORIGIN			
Query Match	20.8%; Score 610.2; DB 10; Length 965;		
Best Local Similarity	85.9%; Pred. No. 1e-113;		
Matches 762; Conservative 0; Mismatches 108; Indels 17; Gaps 7			
OY 1789	ttttgtgcatctttctttacacatttgcgcaaaagtgcagacaaacttggtctcaaa	1848	
Db 4	TTTTGTGGCAATTTCTTTTACACTATTCGCCCAAGGTGAAGTCAAAACCTGGGCTCTCA	63	
OY 1849	tgatcacatgacacagacactgtagaagcaacaataatcctgcgcctctaatgta	1908	
Db 64	TGATACCACTGGACACAGACACTGAAGACCAATTAATTAATTAATTAATTAATTAATTA	123	
OY 1909	ttttattatcatgtaacctcaagctgcctctglatgtgtaagcatctatgcttt	1968	
Db 124	TTTTTATTATTCATGTAACTCAAGTGCGCTCTGTGATGTGTAAACATTTATGTCTTT	183	
OY 1969	ttttaatgtactgttattagatttttaaggcctataatcatgaaataatcaactagtgc	2028	
Db 184	TTTTTATTGTACTTGTATTATTTAGATTTTAAAGGCTTTAAACAATGAATATACACTAGTTC	243	
OY 2029	agaataataaaatgacactgtgtttaataatgaataataatgtaagcttagagacttcaatt	2088	
Db 244	AGAAATTAATTAATTAATTAATCTGTGTTTAAATTAATTAATTAATTAATTAATTAATTA	303	
OY 2089	aggtcacatacctgcctgctagtcgaggcaacatgaagttagagacagttcgttgatttt	2148	
Db 304	AGGTTACATACCTGCTGCTAGTGTGGGGAACATGAAGTAGGACAGTGTGTTGATTTT	363	
OY 2149	tagggcactactaaaggaaatgactgaaacagacctcctgataccttgccttaataa	2208	
Db 364	TAGGCGCAATTAATTAAGGAATGAGCTGAAACAGACCTCCTGTATACCTTGTATTAATA	423	
OY 2209	ctagatgataatctccaggtactgataaaacacctgttgtgttcaacttccctcaataaa	2268	
Db 424	CTAGATGATAATTCCTCAGGTACTGTATAAACACCTGTGTGTATACCTTGTGTCATTAATA	483	

Oy	2269	ttgtcagctctcttcgaaccttaagaacctaaacttagatctcttgga -gcggccac	2327
Dd	484	TTGTAGCTCTCTTCGACACTTAGACCTTAACTTTAGATCTCTGTGGAGGCTGCCTC	543
Oy	2338	cactgtataatttgcgtcgccgcaaccggaactgaagggaagtgtgccagagcgtccaagc	2387
Dd	544	CACGTGTATAATTTCGCCCTGGCAACTGAGACTGAGGGGAAGTGCCCAAGC -TGCCAAGC	602
Oy	2388	actcccctcccggtctcaaggtcagaatgtgccagcggttatc-- agaggcagcatccaa-	2444
Dd	603	ACTCCCTCCCTCGCTTCAGGGTCAGAAGTCCCAAGTTTTAATGAGAAAGCACCATTTCCA	662
Oy	2445	---ggccgagagcagatgtctgaactcttcgctgctgttccttccctc--- gaaggcgctac	2497
Dd	663	GGCCCCAGAGGCGCAATGTGACTCTTTCGCGGGTGCGCTTACTCTTCTGGAAGGACTATC	722
Oy	2498	aatgttgtataaagccctgaagtgaagcaagacaagtgaatccactgtctatgtctgat	2557
Dd	723	AACGTGTATGATTAACGCCCTCGAGTAAGGCGACGAGCAGAACACACGCGCAAGGGCTGG	782
Oy	2558	acatctcctaacttcccttcccccaagcacagagaatatgtctgcacatgaacctgcaaa	2617
Dd	783	AACACTACCACAACACTACTCTTCACACACAGAG -AAAAATGAGGGCATGACCGGGCAA-	839
Oy	2618	agaaaatcgcgaagcgccgcgacgagtggtcattcactgcctgatccccc	2664
Dd	840	--AGAAATGCGAGACCGGGCGCGAGGAGAGTTCAGAGCTCGTAACACAC	884
RESULT	8		
LOCUS	BE867611	676 bp	mRNA linear EST 20-OCT-2000
DEFINITION	601443127P1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3847279 5'		
KEYWORDS	mRNA sequence.		
VERSION	BE867611		
SOURCE	BE867611.1 GI:10316387		
ORGANISM	Homo sapiens		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
COMMENT	1 (bases 1 to 676) NIH-MGC http://mgc.ncl.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://Image.llnl.gov Plate: LImag951 row: e column: 08 High quality sequence stop: 676.		
FEATURES	Location/Qualifiers		
SOURCE	1..676		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3847279"		
	/clone_lib="NIH_MGC_65"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."		
BASE COUNT	145 a 171 c 184 g 176 t		
ORIGIN			

RESULT 10
BG400668 661 bp mRNA linear EST 12-MAR-2001
LOCUS 602464341F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592170 5',
DEFINITION mRNA sequence.
ACCESSION BG400668
VERSION BG400668.1 GI:13294116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 661)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM130 row: b column: 11
High quality sequence start: 7
High quality sequence stop: 656.
Location/Qualifiers
1. 661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592170"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pNNR-LIB (Clontech); Site_1:
SfiI (ggcgctcgcc); Site_2: SfiI (ggcgctatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATCTAGAGGCGCCGAGCGCGCAGTGT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

BASE COUNT 188 a 127 c 122 g 223 t 1 others
ORIGIN

Query Match 19.6%; Score 575.6; DB 10; Length 661;
Best Local Similarity 97.6%; Pred. No. 1.1e-106;
Matches 647; Conservative 0; Mismatches 10; Indels 6; Gaps 6;

QY 1688 attccaggaatggttggtggccgcatcattgctcaaaagctgacccctgatacaactgttggg 1747
Db 1 ATTCCAGGAATGATGGGCGCCGTCATTCGTAAGAGTCGACCCCTGATTAACACTGTTGA 60
QY 1748 gaac-tggcaaacggtgtctatattgctgtgctgctatattgaag-TTTTGGTCCATTTCCTT 1806
Db 61 GAAGTGGCAAAACCGTGTCTATATGCTGCTGCTATTAAG-TTTTGGTCCATTTCCTT 119
QY 1807 tacacattgcgaagaagtgaagtaacaaactgggctcacaatgatacacaatggacacag 1866
Db 120 TACACTATTCGCCAAAGAGTGAAGTACAAACCTGGGCTCTCAATGATGACACCATGGACACAG 179
QY 1867 aacactgaagaaacacataataatctgctcattatgaatgtatttatttattatcatgttaa 1926
Db 180 ACACCTGAAGAACCAATTAATATCCCGCTCATTAATATTTATTTATTCATCATGTAA 239
QY 1927 ccccaagtcgcttcgtatgtgtgaagcattatgtcttttttaattgtactgtat 1986
Db 240 CCTCAAGATGCTTCTGTATGTGTGAAGCATTTATGCTTTTATTAATGTACTGTAT 299

QY 1987 tagattttaaggccataaatacaatgaataatcactagttgcccagaataaataatgaact 2046
Db 300 TAGATTTTAAAGCCCTATATCATGAATAATATCACTAGTTCAGAAATATTAATAATGAAT 359
QY 2047 gtgttaattatgaataataatgtgaagctgggaacttctacttta-ggttcaataactggc 2105
Db 360 GTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 419
QY 2106 tgcctagtcggcaacatgaagtaggaagcttcggtgatttttaaggccatactaa-69 2164
Db 420 TGCTAGTGGGCAACATGAAGTAGACAGTCTGTTGATTTTAAAGGCATCTAAGAG 479
QY 2165 ggaatgagctgaacacagacctcctgataaccttgccttaataactagatgat-aactct 2223
Db 480 GGAATGAGCTGAACACAGACCTCCTGATACCTTGTCTTAATTAACATGAATGAATAATCT 539
QY 2224 caagtcctgaataacacctgtgtgttcaacttctcctaataaatgtcagctcctct 2283
Db 540 CAGGTACTGATTAACACCTGTTGTGTCACTTGTCTCAATTAATAATGTCAG-TCTCTCT 598
QY 2284 gacacttagacctcaacattagacatctctgtgagctgacacactgtataattcgc 2343
Db 599 GACACTTAGACCTCACTTTAGCATCTCTGTGAGCTGCATCCACTGTATTAATTCGG 658
QY 2344 ctg 2346
Db 659 CTG 661

RESULT 11
B1860521 736 bp mRNA linear EST 10-OCT-2001
LOCUS 603386787F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5',
DEFINITION mRNA sequence.
ACCESSION B1860521
VERSION B1860521.1 GI:16001268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM1208 row: m column: 21
High quality sequence stop: 713.
Location/Qualifiers
1. 736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5395772"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6, Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 216 a 142 c 147 g 231 t
ORIGIN

Query Match 19.5%; Score 572.2; DB 10; Length 736;
Best Local Similarity 99.3%; Pred. No. 5,3e-106;
Matches 585; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 620 acttccaaagcgtgtccaaagcattctgtccataaaagttcataatacaacggt 679
|||||
Db 1 ACTTCCAAAGCGGTCCAAAGCAATTCGTCCTCCATTAAGATTCAATCAATCAAGCGGT 60
|||||

QY 680 aagaatgacaaatggatgcagaactcaagaatgattctcgttctctttttttatgac 739
|||||
Db 61 AAGAAATGCCAATGGATGAGAACTCAAGATGGATTCGGTTCCTTTTATAGGC 120
|||||

QY 740 taacatcacacagaattcctcgtgaagatagttgcacagaataaagggggaatactcgt 799
|||||
Db 121 TACATCATCACACAGATTCCTGGAGATATGTTGCCAGCAAAATAGGGGGAATATGCTG 180
|||||

QY 800 ctagaattcggagatcctcgtcagctgtcctcaccctgttcaactccattcgtcagat 859
|||||
Db 181 CTAGGATTTGGGATCCTTGCGACCTGCTGCTCACCCCTTACCTCCATTTGCTGCAGAT 240
|||||

QY 860 ttagagattggaccactcattgtactcagagacactagaagacactagaaggtgttaca 919
|||||
Db 241 TTAGGAGTTGGACCACTATTGTACTCAGACACTAGAAAGACTAGAGAGGGTGTACA 300
|||||

QY 920 ttccagccatgcatgcactgtgtctcttctggctcccccctcttgaagaacaaact 979
|||||
Db 301 TTTCCAGCATGATGCCATGTGTGTTCTTGCGGCTCCCTCTTAAAGAGCAAACTT 360
|||||

QY 980 cttagacattcgtatgcagaagacagcttggagacagtaattctctcctcttctga 1039
|||||
Db 361 CTTAGATTTCAATATCAGAGAGACAGCTTGGGACAGTAAATTTCTCTCTTCTGGA 420
|||||

QY 1040 ataattgtactatataatgattggaattgtcttacttacttcttcttcttcttctga 1099
|||||
Db 421 ATTAATTTGCTACTATATGAAATTTGGACTTATGTTCTTAC-TTTTGTGACTATGTGAATA 479
|||||

QY 1100 ttctggttctcttctgtgcactgtgtatgtagtgcacacacacaaacacagaagaatt 1159
|||||
Db 480 TTTTGGTTCTTTTGTGATCTGTGTAGTTAGTGACACACCAAAACACAAAGACAATT 539
|||||

QY 1160 tcccatatgaaaagaataacattcttcaatcaatgaagaatcagctt 1208
|||||
Db 540 TCCCATTAATGAAAAGAAATACATTTCTTCATCATTAAGAAATCAGGTAT 588
|||||

RESULT 12
BE878141 931 bp mRNA linear EST 20-OCT-2000
LOCUS 601487630F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890071 5',
DEFINITION mRNA sequence.
ACCESSION BE878141
VERSION BE878141.1 GI:10326917
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgi.mgi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DMP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: LHAM9672 row: 1 column: 08
High quality sequence stop: 594.
Location/Qualifiers

source 1..931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890071"
/clone_11b="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pCMV-Sport6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

BASE COUNT 251 a 239 c 276 g 165 t

ORIGIN

Query Match 18.9%; Score 554.6; DB 10; Length 931;
Best Local Similarity 99.0%; Pred. No. 2e-102;
Matches 579; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 2340 tcgacctgcaactgacatgagagagtggtgcacagcagctgccaaagcactccctcctg 2399
|||||
Db 1 TCGCCCTGCAACTGGACTGAGTGAAGGAGTGTGCCAGGCACTGCCAAGCACTCCCTCCTG 60
|||||

QY 2400 gcttcaggtcagagtgcccaagcgttatacagaagcagatcccaagcccaagcagtg 2459
|||||
Db 61 GCTTCAAGGTCAGAGTGCCTCAGGCTTATCAGAGCAGCATCCAAAGCCAGAGCGTGT 120
|||||

QY 2460 cgaactctcgtgtgtccttctcctcctcgtgagggctacaaatgtgtatagaacccctgag 2519
|||||
Db 121 CGACTTTCGGCTGGTGGCTTTCCTCTGAGGGGCTATCAATGTGTGATTAAGCCCTGAG 180
|||||

QY 2520 taggcaagaagatgagatccactgtcactgtatgacttataatcatccctaaacttcccttc 2579
|||||
Db 181 TAGGCAAGACAGTGAATGCCACTGCTATGCTTGTATACATCCTTAACCTTCCCTTCC 240
|||||

QY 2580 cagcacaagagatatctgctcgtgcacatgcactgcacaaagaataatgcagacgagc 2639
|||||
Db 241 CAGCACAGAGGATA-TGCTGGCATGCAACTGCAGAAAGAAATATGCAAGCGG-CGGG 298
|||||

QY 2640 caagtggtcactatgctgtatcccaagcaacttggggggctgagtggtggcaatcatgag 2699
|||||
Db 299 CACGGTGGCTCATGCTGTATATCCAGCACTTGGGGGCTGAGGTGGCATCATGTAG 358
|||||

QY 2700 atcagaagttcagagacacacctgcagcatgtgaaacccactctctactaaataaata 2759
|||||
Db 359 ATCAGGAGTTCCAGACCACTGCGCAGCATGTGAAACCCATCTCTTAATAATATA 418
|||||

QY 2760 aaaaattagcttggtcgtgtgacggtgcgttgaatcccaagatactcagagagctgaggt 2819
|||||
Db 419 AAAAATTAGCTGGGCGTGTGTGACGGGCGCTGTAAATCCAGATTAATCAAGAGGCTGAGGT 478
|||||

QY 2820 agggagatcacttgaacccctggagagtggaagtgcagtggaacccaagatacagccactgca 2879
|||||
Db 479 AGGAGATACATTGAACCTGGAGATGAAGTGTGAGTGAACCAAGATACCGCACCTGCA 538
|||||

QY 2880 ctccagcctcgtgcagatgcagagcactcacaactcaaaaaaataa 2924
|||||
Db 539 CTCACGCTGGGCGATGGAGCGAGACTCCAATCAAAAAAATAA 583
|||||

RESULT 13
BF312727 541 bp mRNA linear EST 21-NOV-2000
LOCUS BF312727
DEFINITION 601898247F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4127794 5',
ACCESSION BF312727
VERSION BF312727.1 GI:11260579
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 541)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: L1CM1019 row: m column: 11
High quality sequence stop: 537.
Location/Qualifiers

FEATURES
source

1. 541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4127794"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(g). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library."

BASE COUNT 153 a 132 c 151 g 105 t
ORIGIN

Query Match 17.9%; Score 524; DB 10; Length 541;
Best Local Similarity 99.8%; Pred. No. 3.4e-96;
Matches 533; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2396 cctgtgctcagggtcagagtgccagcgcttatcatagaggcagcatccaaagccagagcca 2455
|||||
DB 2 CCTGGCTTCAAGGTCACAGATGCCAGCGTTTATAGAGCGACATCCAGCCCGAGACCA 61
QY 2456 gtgtcagactcctcggtcgtgtgcttcctcctcagagggtatcaatgtgtagataagccc 2515
|||||
DB 62 GTGTGCACTCTTCGGCTGCGCTTCTCTCTGAGGGCGTATCAATGTGTAGATTAAGCCC 121
QY 2516 tgaatagaagaagcagtgatgactgctatgtcttgatatactctaacttccc 2575
|||||
DB 122 TGAATAGGCAAGACGATGAGATCAGCTGATGCTTGTATATCATCTCAAACTTTCCC 181
QY 2576 ttcccaagacagaagaaatattgctgtgcatcgaaccctgcaaaagaataatgcgaagcgc 2635
|||||
DB 182 TTCCACACAGAGAGATATTGGCTGGCATGCAACCTGCAAAAGAAAATGCGAAGGGCC 241
QY 2636 cggagcagcgtgtcattcgtcttaatcccaagcacttggggggctgaagtgtggcgaatca 2695
|||||
DB 242 CGGGCAGCGTGGCTCATGCTGTATATCCAGACTTTGGGGGGCTGAGTGGCGGAATCA 301
QY 2696 tgaagataagaattcgagacagccttgccaagatlgltgaaacccactctactataaaa 2755
|||||
DB 302 TGAATAGAGAGTTCGAGACCAAGCTTGCCAGCATGGTGAACCCCAATCTCTATAAAA 361
QY 2756 tacaanaaatatagctggtgctgtgtgaacggcgccctgtaatcccaataaccagaagctg 2815
|||||
DB 362 TACAAAAAATTAGCTGGGGCTGTGTGACGGCGCTGTATATCCAGATACTCAGAGGCTG 421
QY 2816 aggtlaagagaataacttgaacctgtgagagctgtgaagltgcagltgaaccaagaatacagccac 2875
|||||
DB 422 AGGTAGAGAGATCACTTGAACCTGGAGGTGGAAGTTGCAAGTGAACCAAGATCAGCGCAC 481
QY 2876 tgcactcca-gcctgggcgatgtgagcgagactccaactcaaaaaaaaaaaaaa 2930
|||||
DB 482 TGCACTCCAGGCGTGGCGATGAGCGACGACTCCAACTCAAAAAAAAAAAAAA 537

RESULT 14
BI661062 801 bp mRNA linear EST 12-SEP-2001
LOCUS BI661062
DEFINITION 603304362F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350046 5',
mRNA sequence.
ACCESSION BI661062
VERSION BI661062.1 GI:15575298
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 801)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L1AM1889 row: 1 column: 15
High quality sequence stop: 799.
Location/Qualifiers

FEATURES
source

1. 801
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5350046"
/clone_lib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORE; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: oligo dT.
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
Library."

BASE COUNT 173 a 200 c 220 g 208 t
ORIGIN

Query Match 17.9%; Score 523.8; DB 10; Length 801;
Best Local Similarity 81.9%; Pred. No. 3.6e-96;
Matches 641; Conservative 0; Mismatches 137; Indels 5; Gaps 3;

QY 380 ggcgtcatgaggtctccggttcagagacttgcgcggaagaatgagcgaagacagcgc 439
|||||
DB 22 GGGGTTATGAGGCCCTCGTTCGGGGTCCGGCGGAACGACGATGAGAGACTCGGAC 81
QY 440 cgaacgcctcttaccgggcccacagcgccgaagcgcgtccagtgtgtctctgtc 499
|||||
DB 82 AGCAACCCGCTCTGCGGGGCGCCGCGCAAGCGGCTCCAGTGTCTCTCTGCT 141
QY 500 cgttaacttaagaatttggccttttggcttcttcaatgtgtatgattacgtgtg 559
|||||
DB 142 CGGTACAACTTACGATTTTGGCGTTCTGTGTTTCTCTCTATGCTTACGGGTG 201
QY 560 aatctagtgctgcttagtgaatagtagatccaataccaacttaagaataataga 619
|||||
DB 202 AACCTGAGTGTTCGTTAGTGAACATGATGATTTCAATCACTTGACTGATTAATAGA 261
QY 620 acttcaagcgtgtccagagacttctgtccataaagtatcatcataactcaaacggt 679
|||||
DB 262 ACCTCTAAGAGATGTGGGAACATCTGCCCCCATTAAGATTACACACATCAGAGGT 321

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 15, 2002, 09:44:20 ; Search time 592.65 Seconds
(without alignments)
8488.247 Million cell updates/sec

Title: US-09-776-865-1
Perfect score: 2930
Sequence: 1 gtctgcgtcgaagccctccccc.....ctcaaaaaaaaaaaaaaa 2930

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2930	100.0	2930	21	AAZ50879
2	2930	100.0	2930	22	AAZ50879
3	2602	88.8	2602	21	AAZ50875
4	2493.2	85.1	2712	22	AAK94876
5	2359.4	80.5	2512	22	AAZ50875
6	2018.6	68.9	2670	22	AAH79234
7	1570.8	53.6	2844	21	AAZ50876
8	1570.8	53.6	2844	22	AAZ50876
9	1484.8	50.7	1488	22	AAI58115

10	1413	48.2	1485	21	AAZ50880	Human/sheep consen
C 11	1405.6	44.0	1975	22	AAH99626	Human protein enco
12	1305	44.5	1485	21	AAZ50881	Human/sheep consen
13	912.2	31.1	929	22	AAI59901	Human polynucleoti
14	827.8	28.3	853	22	AAK93901	Human cDNA clone r
15	775.6	26.5	838	22	AAK92364	Human cDNA 5'-end
16	541	18.5	1975	22	AAH99626	Human protein enco
C 17	511.2	17.4	535	22	AAH99626	Human cDNA 3'-end
18	434.8	14.8	798	23	AAZ56219	DNA encoding novel
19	434.8	14.8	1066	23	AAZ77186	DNA encoding novel
20	375	12.8	375	22	AAH52158	Human ATP protein
C 21	327.2	11.2	349	22	AAI13566	Human breast cance
22	292	10.0	300	20	AAZ14190	Human gene express
C 23	261.4	8.9	264	22	AAI22435	Human breast cance
C 24	257	8.8	267	22	AAI24574	Human breast cance
25	253.2	8.6	3407	22	AAZ9852	Genomic sequence #
26	253.2	8.6	3407	22	AAZ9852	Human digestive sy
C 27	251	8.6	433	22	AAI24174	Human breast cance
28	250.6	8.6	1992	22	AAK85653	Human immune/haema
29	250.6	8.6	1995	22	AAK85652	Human immune/haema
C 30	249.8	8.5	3422	22	AAK52406	Human polynucleoti
C 31	248.6	8.5	3608	22	AAK67271	Human immune/haema
32	248.6	8.5	3608	22	AAK74891	Human immune/haema
C 33	248.6	8.5	3608	22	AAK83192	Human immune/haema
C 34	248.4	8.5	3178	22	AAK80500	Human immune/haema
C 35	246.8	8.4	3178	22	AAK80499	Human immune/haema
36	246.4	8.4	325791	22	AAZ43104	Human Oestrogen re
37	246.2	8.4	384	22	AAK67754	Human immune/haema
38	246.2	8.4	393	22	AAK67752	Human immune/haema
39	246.2	8.4	393	22	AAK67753	Human immune/haema
40	245.6	8.4	19371	22	AAI05706	Human reproductive
41	245.6	8.4	27154	22	AAI05706	Human reproductive
C 42	245.6	8.4	31952	22	AAK89370	Human digestive sy
C 43	245.6	8.4	48203	22	AAK90161	Human immune/haema
C 44	245.6	8.4	48203	22	AAK81663	Human immune/haema
45	245.6	8.4	48203	22	AAK82628	Human immune/haema

ALIGNMENTS

RESULT 1	
AAZ50879	AAZ50879 standard; cDNA; 2930 BP.
ID	AAZ50879
XX	AAZ50879;
AC	
XX	
DT	31-MAY-2000 (first entry)
XX	
DE	Full length human GBS toxin receptor (HP59) cDNA.
XX	
KW	Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW	pathological vasculature; cancer metastases; angiogenesis;
KW	neovascularisation; reperfusion injury; scarring; keloid;
KW	chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW	endothelial cell proliferation; antibacterial; anticancer;
KW	anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key
FT	Location/Qualifiers
FT	263..1873
FT	/*tag= a
FT	/product= "Human GBS toxin receptor protein"
XX	
PN	W0200005375-A1.
XX	
PD	03-FEB-2000.
XX	
PF	22-JUL-1999; 99WO-US16676.
XX	
PR	22-JUL-1998; 98US-0093843.
XX	

PA (UYVA-) UNIV VANDERBILT.
XX Hellervast CG, Fu C;
XX
DR WPI: 2000-205377/18.
DR P-PSDB; AA145089.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Claim 3; Page 89-93; 109pp; English.
XX
CC The present cDNA sequence encodes full length human GBS (group B beta
CC -haemolytic streptococci) toxin receptor (HP39). This sequence was cloned
CC by using human embryo lung cDNA library as template. Expression vectors
CC comprising this cDNA can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
XX Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

Query Match 100.0%; Score 2930; DB 21; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtcgagagccctcccttaattatgtgcaattcaatcccaatgccgcgcgca 60
DB 1 gtctgtcgagagccctcccttaattatgtgcaattcaatcccaatgccgcgcgca 60
QY 61 gcccccatactcgtcgtcgagaggtgccccctgcaactttaagaagggtgtcagga 120
DB 61 gcccccatactcgtcgtcgagaggtgccccctgcaactttaagaagggtgtcagga 120
QY 121 gcggagagacggtctccgnaacaccggtcccccgcatgtcagacggcgagagcgg 180
DB 121 gcggagagacggtctccgnaacaccggtcccccgcatgtcagacggcgagagcgg 180
QY 181 gctcactttgcgcaatccctacagagaactcccaagaactccgcttcccttaaccacaa 240
DB 181 gctcactttgcgcaatccctacagagaactcccaagaactccgcttcccttaaccacaa 240
QY 241 gccagaattgcccaacctaagatgagcgaggcgagatcacaccgcgcgcggtcca 300
DB 241 gccagaattgcccaacctaagatgagcgaggcgagatcacaccgcgcgcggtcca 300
QY 301 gccagctcggccgggggtctcggtctcggtcggtcggtccctctctctcgcaggtgac 360
DB 301 gccagctcggccgggggtctcggtctcggtcggtcggtccctctctctcgcaggtgac 360
QY 361 gaatacacccgtctacagctgaagcgctcatgaggtctcgttcgagaaactggtccggaaaga 420
DB 361 gaatacacccgtctacagctgaagcgctcatgaggtctcgttcgagaaactggtccggaaaga 420
QY 421 tggcgaagaagacgacgacgcgcctcttaccggtggccccaaggcgcggaagcgcg 480
DB 421 tggcgaagaagacgacgacgcgcctcttaccggtggccccaaggcgcggaagcgcg 480
QY 481 tccaggtgtgtctgtctgtctacaaactagaaatttggccttttttggttcttcat 540
DB 481 tccaggtgtgtctgtctgtctacaaactagaaatttggccttttttggttcttcat 540
QY 541 tgtgtatgcatctagtgtaactgagtggtctgtagtgatagatgtagatcaataac 600
DB 541 tgtgtatgcatctagtgtaactgagtggtctgtagtgatagatgtagatcaataac 600

QY 601 aacttaagaataagaactcccaagcggtgccagagcattctgtcccaataaagt 660
DB 601 aacttaagaataagaactcccaagcggtgccagagcattctgtcccaataaagt 660
QY 661 tcatcataatcaaacggtgaagaagtaaccaatggatgtcagaaactcaagaatgatct 720
DB 661 tcatcataatcaaacggtgaagaagtaaccaatggatgtcagaaactcaagaatgatct 720
QY 721 cgggtcctttttatggcacaatccacaacagattcccgagagatgttgcggca 780
DB 721 cgggtcctttttatggcacaatccacaacagattcccgagagatgttgcggca 780
QY 781 aatagggggaaaatctgctgaagatttggatcccttgcacgcgtccctccctgt 840
DB 781 aatagggggaaaatctgctgaagatttggatcccttgcacgcgtccctccctgt 840
QY 841 caatccattgtgcagatttaagagltggacaatcaatgttaactcaagagactagaag 900
DB 841 caatccattgtgcagatttaagagltggacaatcaatgttaactcaagagactagaag 900
QY 901 actagagaggggttaccatttccacatgcatgcatgtgtcttcttggtctccccc 960
DB 901 actagagaggggttaccatttccacatgcatgcatgtgtcttcttggtctccccc 960
QY 961 tcttgaagaagacaactcttagcatttcgtatgcagagacagcttggagcaat 1020
DB 961 tcttgaagaagacaactcttagcatttcgtatgcagagacagcttggagcaat 1020
QY 1021 ttctctcctctctctcgtgaataattgtctaataatgtgaatttgacttaagtcttact 1080
DB 1021 ttctctcctctctctcgtgaataattgtctaataatgtgaatttgacttaagtcttact 1080
QY 1081 ttgtgtactatggatatttgggttcttcttggtgattctgtgttagtgaacacc 1140
DB 1081 ttgtgtactatggatatttgggttcttcttggtgattctgtgttagtgaacacc 1140
QY 1141 acaaaaacacaagagatcccatatgaagaagatacatcttcatcatgaaga 1200
DB 1141 acaaaaacacaagagatcccatatgaagaagatacatcttcatcatgaaga 1200
QY 1201 tcagcttcttcaacagaagtcagtcggtggtgtaccatlttaaaatccctgcacttgg 1260
DB 1201 tcagcttcttcaacagaagtcagtcggtggtgtaccatlttaaaatccctgcacttgg 1260
QY 1261 ggtcatcgtatgtgcacacttcttacaacttgacttttaacttttagacttatt 1320
DB 1261 ggtcatcgtatgtgcacacttcttacaacttgacttttaacttttagacttatt 1320
QY 1321 gcttactatatagaagagatccctaaggttcaatgtctcaagaagaatgttttacttc 1380
DB 1321 gcttactatatagaagagatccctaaggttcaatgtctcaagaagaatgttttacttc 1380
QY 1381 atgtccttatttaggcctctgtgtatgtatgatacctgtctgtcaagctgtgaacaatt 1440
DB 1381 atgtccttatttaggcctctgtgtatgtatgatacctgtctgtcaagctgtgaacaatt 1440
QY 1441 aagggcaaaatggaatttttcaactttagtgttcgaagaatttttagccttaagaat 1500
DB 1441 aagggcaaaatggaatttttcaactttagtgttcgaagaatttttagccttaagaat 1500
QY 1501 gattggaccctgcagatcttcctgttagctgtggtcctcaatgtgtgatatctcttgc 1560
DB 1501 gattggaccctgcagatcttcctgttagctgtggtcctcaatgtgtgatatctcttgc 1560
QY 1561 cgttgccttccctaataatacaaacacgtggagagcttctgtcctcgtatagcat 1620
DB 1561 cgttgccttccctaataatacaaacacgtggagagcttctgtcctcgtatagcat 1620
QY 1621 caaccatctggatatgtccctctgtatgctgttatccctcgggggatacaacaatacat 1680
DB 1621 caaccatctggatatgtccctctgtatgctgttatccctcgggggatacaacaatacat 1680

QY 1681 tggcaatattccagaatgtgttgccgctcattgtctaaagtctgaccctgataaac 1740
 Db 1681 tggcaatattccagaatgtgttgccgctcattgtctaaagtctgaccctgataaac 1740
 QY 1741 tgttggagaatggaacacggtgtctatatgtctgtctatnaatgttttgggtccat 1800
 Db 1741 tgttggagaatggaacacggtgtctatatgtctgtctatnaatgttttgggtccat 1800
 QY 1801 ttctcttaactcttgcgcaaaagtgaatgacaacatggtggtcctaataatgatacatag 1860
 Db 1801 ttctcttaactcttgcgcaaaagtgaatgacaacatggtggtcctaataatgatacatag 1860
 QY 1861 acacagacactggaaggacaataataatcctgctcctatnaatgtattttattatc 1920
 Db 1861 acacagacactggaaggacaataataatcctgctcctatnaatgtattttattatc 1920
 QY 1921 atgtaacctccaagtgcctctgtatgtgtgaagcattctatgtcttttttaattgtac 1980
 Db 1921 atgtaacctccaagtgcctctgtatgtgtgaagcattctatgtcttttttaattgtac 1980
 QY 1981 ttgtattagatttttaaggcctaatactgaataatcctagttgcaggaataataaa 2040
 Db 1981 ttgtattagatttttaaggcctaatactgaataatcctagttgcaggaataataaa 2040
 QY 2041 tgaactgtgtttaataatgaataataatgaagctaggaacttctactttagttcacatac 2100
 Db 2041 tgaactgtgtttaataatgaataataatgaagctaggaacttctactttagttcacatac 2100
 QY 2101 ctgctcgtcagtcggtggaacatgaagtaagaacgtctgtttgatttttaaggcact 2160
 Db 2101 ctgctcgtcagtcggtggaacatgaagtaagaacgtctgtttgatttttaaggcact 2160
 QY 2161 aaagggaatgagctggaacagacccctgataccttgccttaataacagatgataat 2220
 Db 2161 aaagggaatgagctggaacagacccctgataccttgccttaataacagatgataat 2220
 QY 2221 tctcaggtaactgaataaacctgtgtgttctcacttccatcaaaaatgttcaagctcc 2280
 Db 2221 tctcaggtaactgaataaacctgtgtgttctcacttccatcaaaaatgttcaagctcc 2280
 QY 2281 tctgaactatgaaccccaactttagcatcctgtgtgaagctgcacactgtataatt 2340
 Db 2281 tctgaactatgaaccccaactttagcatcctgtgtgaagctgcacactgtataatt 2340
 QY 2341 cgccttgcaacttgacttgaaggaggtgtgcccagagcgtgcgcaagcactccctccctg 2400
 Db 2341 cgccttgcaacttgacttgaaggaggtgtgcccagagcgtgcgcaagcactccctccctg 2400
 QY 2401 ctccaaggtcagaatgcccagcgtttatcagaagcagcatccaagccagagccagttgc 2460
 Db 2401 ctccaaggtcagaatgcccagcgtttatcagaagcagcatccaagccagagccagttgc 2460
 QY 2461 gactcttggcgtgtgctcttctcctcgaagggtatcaatgtgtagataaaagccctgagt 2520
 Db 2461 gactcttggcgtgtgctcttctcctcgaagggtatcaatgtgtagataaaagccctgagt 2520
 QY 2521 aggcagaagcagatgagatcactgctatgtcttgaatacctcctaacttccctcc 2580
 Db 2521 aggcagaagcagatgagatcactgctatgtcttgaatacctcctaacttccctcc 2580
 QY 2581 agcacagagaatattgtgctgtgcatgcaacctgcaaaaagaanaatgcaagcggccgggc 2640
 Db 2581 agcacagagaatattgtgctgtgcatgcaacctgcaaaaagaanaatgcaagcggccgggc 2640
 QY 2641 acgggtgtccatgtcctgttaatcccaactcttggtgggtgtgagtggtgcaatcatgaga 2700
 Db 2641 acgggtgtccatgtcctgttaatcccaactcttggtgggtgtgagtggtgcaatcatgaga 2700
 QY 2701 tcagaggttcgagacagcctggtccagcatgtgtgaaccccatcttactaataataacaa 2760
 Db 2701 tcagaggttcgagacagcctggtccagcatgtgtgaaccccatcttactaataataacaa 2760
 QY 2761 aaataagctggcgtgtgtgaacggcgctgttaatccagatactcagaggtctgagta 2820

Db 2761 aaataagctggcgtgtgtgaacggcgctgttaatccagatactcagaggtctgagta 2820
 QY 2821 ggagaatcaactggaacctggtggaggtgtgagttgcagttgacaagaatcagccactgac 2880
 Db 2821 ggagaatcaactggaacctggtggaggtgtgagttgcagttgacaagaatcagccactgac 2880
 QY 2881 tccagcctggtgagtgagcagagactccactcaacaaataaaaaaa 2930
 Db 2881 tccagcctggtgagtgagcagagactccactcaacaaataaaaaaa 2930

RESULT 2

AADI0325

ID AADI0325 standard; DNA; 2930 BP.

AC AADI0325;

DT 16-OCT-2001 (first entry)

DE Human group B beta-haemolytic Streptococci toxin receptor (HP59) DNA.

KW Human; group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;

KW cytosolic; vulnery; antiatherosclerotic; osteopathic; vasotropic;

KW prevention; attenuation; pathoangiogenic condition; cancer; scar;

KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;

KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;

KW vaccine; ds.

OS Homo sapiens.

PN WO200156598-A2.

PD 09-AUG-2001.

PF 02-FEB-2001; 2001WO-US03662.

PR 02-FEB-2000; 2000US-0179870.

XX (UYVA-) UNIV VANDERBILT.

PI Helligqvist CG;

DR WPI; 2001-48844/53.

XX P-PSDB; AAE06518.

PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,

PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by

PT administering group B beta-hemolytic Streptococci toxin receptor or its

PS fragment -

XX Disclosure: Page 41-44; 52pp; English.

XX The present sequence is a DNA encoding full length human group B beta-

XX haemolytic Streptococci (GBS) toxin receptor protein, HP59. The present

XX invention relates to a method for preventing or attenuating a patho-

XX angiogenic condition in a mammal which comprises administering to the

XX mammal one or more GBS toxin receptors or their immunogenic fragments

XX to induce or maintain an immune response to one of GBS toxin receptors.

XX The method is useful for preventing or ameliorating pathoangiogenic

XX conditions such as cancer, scarring during wound healing, gliosis

XX during repair of nerve injury, chronic wounds, keloids, reperfusion

XX injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and

XX psoriasis in mammals. The proteins of the invention are also used

XX as vaccines.

XX Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

Query Match 100.0%; Score 2930; DB 22; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgttcgaagccctcccttaatttatgtgcaattcaagtcaccaactgcccgcgcga 60
Db 1 gtctgttcgaagccctcccttaatttatgtgcaattcaagtcaccaactgcccgcgcga 60
QY 61 gcccacatactctgtctgtgagcggtgtgcccctgcaatttaagaagggtgcagga 120
Db 61 gcccacatactctgtctgtgagcggtgtgcccctgcaatttaagaagggtgcagga 120
QY 121 gcggagaaaggttcgtccgaacaacggctcccggcatgtgagacggcgaggagcg 180
Db 121 gcggagaaaggttcgtccgaacaacggctcccggcatgtgagacggcgaggagcg 180
QY 181 gctcaatttgcgaactcctcgaagaactccagaactccgttccttaagtccaacca 240
Db 181 gctcaatttgcgaactcctcgaagaactccagaactccgttccttaagtccaacca 240
QY 241 gccgaagttgccacataagatgagcggggagatgaacacggccggccggtcca 300
Db 241 gccgaagttgccacataagatgagcggggagatgaacacggccggccggtcca 300
QY 301 gccagctcgcccgggggtctcgagctgtcgggcgcgctcccttctcgtccaggtgc 360
Db 301 gccagctcgcccgggggtctcgagctgtcgggcgcgctcccttctcgtccaggtgc 360
QY 361 gattacaacgtctacgttaagcggtcattgaggtctccggttcgagaactgtgcggaa 420
Db 361 gattacaacgtctacgttaagcggtcattgaggtctccggttcgagaactgtgcggaa 420
QY 421 tggcgagagagcgacggacggacggctctctacggcgcccgacggcgagagcg 480
Db 421 tggcgagagagcgacggacggacggctctctacggcgcccgacggcgagagcg 480
QY 481 tccagtgctgctgtctcggttaacacttaagcaatttgcccttttggttcttc 540
Db 481 tccagtgctgctgtctcggttaacacttaagcaatttgcccttttggttcttc 540
QY 541 tgtgtatgcatatgctgtgaatcgtgaatgttgcgttagtgatagatagatcaaat 600
Db 541 tgtgtatgcatatgctgtgaatcgtgaatgttgcgttagtgatagatagatcaaat 600
QY 601 aacttgaagataatagaactccaagcggtgccagagacattctcccataaagt 660
Db 601 aacttgaagataatagaactccaagcggtgccagagacattctcccataaagt 660
QY 661 tcatcataatcaaacgggttaagaagtacaaatgggattgcagaaaatcaagaatg 720
Db 661 tcatcataatcaaacgggttaagaagtacaaatgggattgcagaaaatcaagaatg 720
QY 721 cgggtctttttttatggtacatcatcacagaatcccgaggagatgttgcgaaca 780
Db 721 cgggtctttttttatggtacatcacagaatcccgaggagatgttgcgaaca 780
QY 781 aataggggggaaaaatgtcttagagatttggaatcctgtgcactgtctccaccgt 840
Db 781 aataggggggaaaaatgtcttagagatttggaatcctgtgcactgtctccaccgt 840
QY 841 cactcccatgtgcagatattagagatttggaacatcatgtaactcagagactaaga 900
Db 841 cactcccatgtgcagatattagagatttggaacatcatgtaactcagagactaaga 900
QY 901 actagagagaggtgttaacattccagcagcagatgtggtcttcttgggtctcccc 960
Db 901 actagagagaggtgttaacattccagcagcagatgtggtcttcttgggtctcccc 960
QY 961 tcttgaagaagaacaactcttagcatcttctgtacagagcagaactgtggagcaat 1020
Db 961 tcttgaagaagaacaactcttagcatcttctgtacagagcagaactgtggagcaat 1020

QY 1021 ttctctctctctcttggaaatattgtactatgatgaattggaattgtctctact 1080
Db 1021 ttctctctctctcttggaaatattgtactatgatgaattggaattgtctctact 1080
QY 1081 ttctgttcaatttggaaattttgttcttcttgttgatctgtttagtgaacacc 1140
Db 1081 ttctgttcaatttggaaattttgttcttcttgttgatctgtttagtgaacacc 1140
QY 1141 acaaaaacagaagaatttcccatatgaagaagaatacttctcatcataaga 1200
Db 1141 acaaaaacagaagaatttcccatatgaagaagaatacttctcatcataaga 1200
QY 1201 tcaagcttcttccagaagaatgcgttgggttaaccatttlaaataccctgcact 1260
Db 1201 tcaagcttcttccagaagaatgcgttgggttaaccatttlaaataccctgcact 1260
QY 1261 ggctatcgtatgtgcacatttctcttaaacctggaacttttaacttaattga 1320
Db 1261 ggctatcgtatgtgcacatttctcttaaacctggaacttttaacttaattga 1320
QY 1321 gctcaattatgaagaagatcctcaaggttcaaatgttcaagagaatgtttatct 1380
Db 1321 gctcaattatgaagaagatcctcaaggttcaaatgttcaagagaatgtttatct 1380
QY 1381 atgtccttatttaggctctgtgtatgtatgatccctgtcgtcaagctgtgaca 1440
Db 1381 atgtccttatttaggctctgtgtatgtatgatccctgtcgtcgaagctgtgaca 1440
QY 1441 aagggcgaatatttccaacttatgttctcgagaatttttagccttatagaat 1500
Db 1441 aagggcgaatatttccaacttatgttctcgagaatttttagccttatagaat 1500
QY 1501 gattggacccgcagatctctgttagctgtgtctgtcatgtggtgatatcttgg 1560
Db 1501 gattggacccgcagatctctgttagctgtgtctgtcatgtggtgatatcttgg 1560
QY 1561 cgtgtcttctcaactatcaacaacactggagagcttctgtctctgtgattag 1620
Db 1561 cgtgtcttctcaactatcaacaacactggagagcttctgtctctgtgattag 1620
QY 1621 caaccatctgatatgtctctctgtatgttgttatctctctgtggacataatcat 1680
Db 1621 caaccatctgatatgtctctctgtatgttgttatctctctgtggacataatcat 1680
QY 1681 tgcacactatccagaatgttggccggtcatgtgctaaagctgaacccctgataac 1740
Db 1681 tgcacactatccagaatgttggccggtcatgtgctaaagctgaacccctgataac 1740
QY 1741 tgttggagaatggcaaacgtgttctatattgtctgtctatgaatgttttgg 1800
Db 1741 tgttggagaatggcaaacgtgttctatattgtctgtctatgaatgttttgg 1800
QY 1801 ttcttcaacatacttgcgaagaagtgaagaacaacatgggctcctaagtacaatg 1860
Db 1801 ttcttcaacatacttgcgaagaagtgaagaacaacatgggctcctaagtacaatg 1860
QY 1861 acacagacactggaagaacaaataatccgcctcttaattgaattttatttacc 1920
Db 1861 acacagacactggaagaacaaataatccgcctcttaattgaattttatttacc 1920
QY 1921 atgttaacctaaagtgcctctgtatgttgaagcatctcatalgtctttttaa 1980
Db 1921 atgttaacctaaagtgcctctgtatgttgaagcatctcatalgtctttttaa 1980
QY 1981 ttgtattagaattttaaaggcttaactatgaataatcaactagttcgagaataaaa 2040
Db 1981 ttgtattagaattttaaaggcttaactatgaataatcaactagttcgagaataaaa 2040
QY 2041 tgaactgttctaattatgaatataatgaagcttaagacttcaactttagttcacat 2100
Db 2041 tgaactgttctaattatgaatataatgaagcttaagacttcaactttagttcacat 2100
QY 2101 ctgctgtctagtcggcacaatgaagtaaggaacgtctgttatttttaggcata 2160

```

DB 2101 ctcgcccgtacgtcgccgcaacatgaagtagacagtcctgttatttttagggccatatt 2160
OY 2161 aaaggaatgaagctggaacagacccctccgtatagcttctgttaataatagatgaat 2220
DB 2161 aaaggaatgaagctggaacagacccctccgtatagcttctgttaataatagatgaat 2220
OY 2221 tctcaggtactataaacacccgtgtgttctacttctccataaaatgtcagctcc 2280
DB 2221 tctcaggtactataaacacccgtgtgttctacttctccataaaatgtcagctcc 2280
OY 2281 tctgcaacttagaacctcaaaactttagcatctctgtgtgagctgcacatcgtataatt 2340
DB 2281 tctgcaacttagaacctcaaaactttagcatctctgtgtgagctgcacatcgtataatt 2340
OY 2341 gcgcctggcgaactggaactgagggagtggtccagagcagctgcacacccctccctgg 2400
DB 2341 gcgcctggcgaactggaactgagggagtggtccagagcagctgcacacccctccctgg 2400
OY 2401 ctccagagtcacagatgcccagcgtttatcacagagcagcatcccaagccagagccagtgc 2460
DB 2401 ctccagagtcacagatgcccagcgtttatcacagagcagcatcccaagccagagccagtgc 2460
OY 2461 gactcttcggtgtgtgcttctcctctgagggctatcaatgtgtagataaagccctgagt 2520
DB 2461 gactcttcggtgtgtgcttctcctctgaggggctatcaatgtgtagataaagccctgagt 2520
OY 2521 aaggcagaagcagatgagatccactgctatgtcttgaataatctcaaaatttccctccc 2580
DB 2521 aaggcagaagcagatgagatccactgctatgtcttgaataatctcaaaatttccctccc 2580
OY 2581 agcacagaagaatattgtgtgtgcacacccctgcacaaagaataatgcgagcgccgggc 2640
DB 2581 agcacagaagaatattgtgtgtgcacacccctgcacaaagaataatgcgagcgccgggc 2640
OY 2641 acggtgtgtcattgctgttaatcccaagcacttggggggtgtgtgtgtgtgtgtgtgtgt 2700
DB 2641 acggtgtgtcattgctgttaatcccaagcacttggggggtgtgtgtgtgtgtgtgtgtgt 2700
OY 2701 tgaagagttcgaagacagcagctgcccagcagctgtgtgaacccctctctactaaataaaca 2760
DB 2701 tgaagagttcgaagacagcagctgcccagcagctgtgtgaacccctctctactaaataaaca 2760
OY 2761 aaatagatctggagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2820
DB 2761 aaatagatctggagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2820
OY 2821 ggaagaatcacttgaaccttggaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880
DB 2821 ggaagaatcacttgaaccttggaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880
OY 2881 tccagcgttggcgtatgtagcagagactcccaactcaaaaaaaataaaaaa 2930
DB 2881 tccagcgttggcgtatgtagcagagactcccaactcaaaaaaaataaaaaa 2930

```

RESULT 3

AAZ50875
ID AAZ50875 standard; cDNA; 2602 BP.

AAZ50875;

31-MAY-2000 (first entry)

Partial human GBS toxin receptor (HP55) cDNA.

Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
pathological vasculatisation; cancer metastases; angioneurosis;
neovascularisation; reperfusion injury; scarring; keloid;
chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
endothelial cell proliferation; antibacterial; anticancer;
anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.

```

OS Homo sapiens.
XX Key Location/Qualifiers
XX CDS 58..1545
XX FT /tag=a
XX FT /product="Partial human GBS toxin receptor"
XX PN WO200005375-A1.
XX PD 03-FEB-2000.
XX PF 22-JUL-1999; 99WO-0516676.
XX PR 22-JUL-1998; 98US-0093843.
XX PA (UUYA-) UNIV VANDERBILT.
XX PI Helleqvist CG, Fu C;
XX DR WPI, 2000-205377/18.
XX DR P-PSDB; AAY45087.
XX PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX PS useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX SQ Claim 3; Page 77-80; 109pp; English.

CC The present cDNA sequence encodes partial human GBS (group B beta
CC -haemolytic streptococci) toxin receptor (HP55). This sequence was cloned
CC by using human embryo lung cDNA library as template. Expression vectors
CC comprising this cDNA can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vasculatisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX SQ Sequence 2602 BP; 672 A; 576 C; 583 G; 771 T; 0 other;

Query Match 88.8%; Score 2602; DB 21; Length 2602;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 329 tcgggcggcgctccctctctctgcccaggtggcgaagatccctgctacgttagcgctatg 388
DB 1 tcgggcggcgctccctctctctgcccaggtggcgaagatccctgctacgttagcgctatg 60
OY 389 aggtctccggttcgaagacctggcccgaaagatgagcgaagagcagcagcgcct 448
DB 61 aggtctccggttcgaagacctggcccgaaagatgagcgaagagcagcagcgcct 120
OY 449 ctctaccggcgccccaaggcgccgaagcgcgttccaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 508
DB 121 ctctaccggcgccccaaggcgccgaagcgcgttccaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
OY 509 ttagaacttggccttttttggttctctcatctgtgtatgcatatgtagaactgagt 568
DB 181 ttagaacttggccttttttggttctctcatctgtgtatgcatatgtagaactgagt 240
OY 569 gtgtcgttagtgatagtgtagatccaatacaacttagaagaataaagactccaag 628
DB 241 gtgtcgttagtgatagtgtagatccaatacaacttagaagaataaagactccaag 300
OY 629 ggtgtgtccgaagatctgtctccataaagttcatcataatcaaacgggtagaagatc 688
DB 301 ggtgtgtccgaagatctgtctccataaagttcatcataatcaaacgggtagaagatc 360

```

QY 689 caatggatgcagagaactcaaggaatgattctcgtctctctttttttagtgatacatc 748
|||||
Db 361 caatggatgcagagaactcaaggaatgattctcgtctctctttttttagtgatacatc 420
QY 749 acaagaattctctggaagatatgttgcagcaaaaataagggggaaatgctgctagatt 808
|||||
Db 421 acaagaattctctggaagatatgttgcagcaaaaataagggggaaatgctgctagatt 480
QY 809 gggatctctgcaactgctgcctcctcaccctgttcaactccattgtgtgagattgaggt 868
|||||
Db 481 gggatctctgcaactgctgcctcctcaccctgttcaactccattgtgtgagattgaggt 540
QY 869 ggaacctcatctgactcagagcaactagaaaggagaggggtgtacattctccagcc 928
|||||
Db 541 ggaacctcatctgactcagagcaactagaaaggagaggggtgtacattctccagcc 600
QY 929 atgcatgcattgtgtctctcttggcctccctctgaaagaagcaactcttagcatt 988
|||||
Db 601 atgcatgcattgtgtctctcttggcctccctctgaaagaagcaactcttagcatt 660
QY 989 tcgtatgcagagacagagcttggagacagtaattctctctctcttcttggaataattgc 1048
|||||
Db 661 tcgtatgcagagacagagcttggagacagtaattctctctctcttcttggaataattgc 720
QY 1049 tactataatgaaatgaaatctgtctcttacttttttgtaactatggaataattgttt 1108
|||||
Db 721 tactataatgaaatgaaatctgtctcttacttttttgtaactatggaataattgttt 780
QY 1109 cttttgtgagatctgttgaattagtgacacacacacacacacacacacacacacacac 1168
|||||
Db 781 cttttgtgagatctgttgaattagtgacacacacacacacacacacacacacacacac 840
QY 1169 gaaaaggaaatacaattcttcaactcaactcaactcaactcaactcaactcaactcaact 1228
|||||
Db 841 gaaaaggaaatacaattcttcaactcaactcaactcaactcaactcaactcaactcaact 900
QY 1229 tgggtaccacatttaaaatccctgcagcattgggctatcgatgtgcacacattctctac 1288
Db 901 tgggtaccacatttaaaatccctgcagcattgggctatcgatgtgcacacattctctac 960
QY 1289 aactggacttttactattatgacattatgacttacttataatgaaaggagatccttagg 1348
|||||
Db 961 aactggacttttactattatgacattatgacttacttataatgaaaggagatccttagg 1020
QY 1349 ttcaattctcaagaagaatgggttttattcttcaactgcttatttaaggctctgtttatgt 1408
|||||
Db 1021 ttcaattctcaagaagaatgggttttattcttcaactgcttatttaaggctctgtttatgt 1080
QY 1409 atgactcctgtcgttcaagctgtctgacacatttaagggcaaaatggaattttcaactta 1468
Db 1081 atgactcctgtcgttcaagctgtctgacacatttaagggcaaaatggaattttcaactta 1140
QY 1469 tgggttcagagaatttttaagccttatagaaatgattggaactggaagatcttcctgtatgt 1528
|||||
Db 1141 tgggttcagagaatttttaagccttatagaaatgattggaactggaagatcttcctgtatgt 1200
QY 1529 gctggtcctatctgctgtgattattcttggcgttgcttcttcccaactatatacaaca 1588
|||||
Db 1201 gctggtcctatctgctgtgattattcttggcgttgcttcttcccaactatatacaaca 1260
QY 1589 ctggagagctttgctctcttctgatttagatacaacacatctgatatgtctctgtat 1648
|||||
Db 1261 ctggagagctttgctctcttctgatttagatacaacacatctgatatgtctctgtat 1320
QY 1649 gctggtatctctctgtgcatcacaataacatttgcacatttccaggaatggttggccc 1708
Db 1321 gctggtatctctctgtgcatcacaataacatttgcacatttccaggaatggttggccc 1380
QY 1709 gtaattgctaaagctgcagccctgataacacgttggagaatgccaacgtttgtat 1768
|||||
Db 1381 gtaattgctaaagctgcagccctgataacacgttggagaatgccaacgtttgtat 1440
QY 1769 atgtgtcgtctatlaattgttttgggtgcatttcttcaactatcgcacaaagtgcga 1828

Db 1441 atgtgtcgtctatlaattgttttgggtgcatttcttcaactatcgcacaaagtgcga 1500
QY 1829 gtacaaaactgggctcccaatgatcaccatgagacagacacacacacacacacacacac 1888
Db 1501 gtacaaaactgggctcccaatgatcaccatgagacagacacacacacacacacacacac 1560
QY 1889 atctgcctctatlaattgttttatttcaatgatacctcaacagtgctctgtatgt 1948
Db 1561 atctgcctctatlaattgttttatttcaatgatacctcaacagtgctctgtatgt 1620
QY 1949 tgtgaagatctcatgctctttttttaaattgtactgtatattagattttaaggcctataatc 2008
Db 1621 tgtgaagatctcatgctctttttttaaattgtactgtatattagattttaaggcctataatc 1680
QY 2009 atggaataatacactagtttgcagaaataataaataaataaataaataaataaataaata 2068
Db 1681 atggaataatacactagtttgcagaaataataaataaataaataaataaataaataaata 1740
QY 2069 taagctaggaacttactattaggttcaacataccctgctgtcgtgcggcaacatgaaata 2128
Db 1741 taagctaggaacttactattaggttcaacataccctgctgtcgtgcggcaacatgaaata 1800
QY 2129 ggaacgtctctgtgattttttaagggcatactaaagggaatgagctgcgaacacagactcct 2188
Db 1801 ggaacgtctctgtgattttttaagggcatactaaagggaatgagctgcgaacacagactcct 1860
QY 2189 gataccttctgcttaataataactagatataatctcagagtaacacacacacacacacac 2248
Db 1861 gataccttctgcttaataataactagatataatctcagagtaacacacacacacacacac 1920
QY 2249 gttaacttctcctataaanaattgtcagctctcctcctgcacattgacactcaactttagca 2308
Db 1921 gttaacttctcctataaanaattgtcagctctcctcctgcacattgacactcaactttagca 1980
QY 2309 tctctgtgagacgtgcacatccactgtataatttcgctgtgcacactggaatgagggagatgt 2368
Db 1981 tctctgtgagacgtgcacatccactgtataatttcgctgtgcacactggaatgagggagatgt 2040
QY 2369 gcccaagcagctgcagacacactccctcctgcgtctcagagtgcaagagtgccacagctttat 2428
Db 2041 gcccaagcagctgcagacacactccctcctgcgtctcagagtgcaagagtgccacagctttat 2100
QY 2429 cagagagcagatcccaagcccaagagccagtgctgacttctgcgtgtgtcccttctctga 2488
Db 2101 cagagagcagatcccaagcccaagagccagtgctgacttctgcgtgtgtcccttctctga 2160
QY 2489 ggggctatcaatgtgtatgaataagccctgagtgagcaagagcagtgatccactgtat 2548
Db 2161 ggggctatcaatgtgtatgaataagccctgagtgagcaagagcagtgatccactgtat 2220
QY 2549 ggtctgtatcatcctcctcaaaacttccctcccaagcaagaaggaatattgtctgtgcatga 2608
Db 2221 ggtctgtatcatcctcctcaaaacttccctcccaagcaagaaggaatattgtctgtgcatga 2280
QY 2609 acctgcacaaagaanaatgycgaagcggcggcggcggcggcggcggcggcggcggcggcggc 2668
Db 2281 acctgcacaaagaanaatgycgaagcggcggcggcggcggcggcggcggcggcggcggcggc 2340
QY 2669 ctttggggggctgaggttggcgaatcagatcaggaagtctgagacacagcctggcagc 2728
Db 2341 ctttggggggctgaggttggcgaatcagatcaggaagtctgagacacagcctggcagc 2400
QY 2729 atgtgtgaaccctcatctctataaataataaataaataaataaataaataaataaataa 2788
Db 2401 atgtgtgaaccctcatctctataaataataaataaataaataaataaataaataaataa 2460
QY 2789 ctgtaatcccaagatactcagagagcgtgaagtgagagaatacacttgaacccggagagtgga 2848
Db 2461 ctgtaatcccaagatactcagagagcgtgaagtgagagaatacacttgaacccggagagtgga 2520
QY 2849 agttgcagtgaaacacagatccagccactgacatccagcctgtggcgaatggaagcagactcc 2908
|||||

Db 2521 agtgcagtgtaaccgaatcacgccactgcactccagcttggtgcatgtgagcgagactcc 2580
QY 2909 aactcaaaaaaaaaaaaaa 2930
Db 2581 aactcaaaaaaaaaaaaaa 2602

RESULT 4
AAK94876
ID AAK94876 standard; cDNA; 2712 BP.
XX
AC AAK94876;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 4068.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EPI130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
XX
P-PSDB: AAM93914.
XX

830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 4068; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
CC
XX
XX Sequence 2712 BP; 677 A; 619 C; 619 G; 797 T; 0 other;

Query Match 85.1%; Score 2493.2; DB 22; Length 2712;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 2607; Conservative 0; Mismatches 3; Indels 102; Gaps 1;

QY 304 agtcgccccgggggtctcggtctgcggcgccgctccctctctgcgaagtgcgag 363
Db 1 agtcgccccgggggtctcggtctgcggcgccgctccctctctgcgaagtgcgag 60
QY 364 taacactgctcacgtatgagctcctccggttcgagacgtgcgcggaacatag 423
Db 61 taacactgctcacgtatgagctcctccggttcgagacgtgcgcggaacatag 120
QY 424 cgaagagagacgacgacgacgctctctaccggcgcccaaggccgaagccgctcc 483

Db 121 cgagagagacgacgacgacgacgctctctaccggcgcccaaggccgaagccgctcc 180
QY 484 agtgcgtctctctgcctcgtttacaacttaggaatttggccttttggttctctatgt 543
Db 181 agtgcgtctctctgcctcgtttacaacttaggaatttggccttttggttctctatgt 240
QY 544 gtaagcattacgtatgaatcgtatgttgcgttagtgatgtatgtatcaatcaac 603
Db 241 gtaagcattacgtatgaatcgtatgttgcgttagtgatgtatgtatcaatcaac 300
QY 604 tttagaagaatagaacttccaaggcgtgtccagagacattcgtccccaataaagttca 663
Db 301 tttagaagaatagaacttccaaggcgtgtccagagacattcgtccccaataaagttca 360
QY 664 tcaatacaacggttaagaagatgacaaatggggtatgagaacatccaagatgattctcg 723
Db 361 tcaatacaacggttaagaagatgacaaatggggtatgagaacatccaagatgattctcg 420
QY 724 ttccctttttaggtacatcatcacagatctcctggaggaatgtgtccagcaaat 783
Db 421 ttccctttttaggtacatcatcacagatctcctggaggaatgtgtccagcaaat 480
QY 784 aggggggaaatgctgtaggaattgggaccccttggaactgctcctcaaccctgtcac 843
Db 481 aggggggaaatgctgtaggaattgggaccccttggaactgctcctcaaccctgtcac 540
QY 844 tcccatgtctgcagatttggagattggagaccccatctgtactaagagcaataaagact 903
Db 541 tcccatgtctgcagatttggagattggagaccccatctgtactaagagcaataaagact 600
QY 904 aggaagggtgttacatttccacagcatgcatgcatgctctcttggtgagcccccct 963
Db 601 aggaagggtgttacatttccacagcatgcatgcatgctctcttggtgagcccccct 660
QY 964 tgaagaagaacaaatctcttagcatcttcgtatgaggaagacagcttgggacaglaattc 1023
Db 661 tgaagaagaacaaatctcttagcatcttcgtatgaggaagacagcttgggacaglaattc 720
QY 1024 tcttcctcttctggaataattgctcactatagaaattgacttatgtcttactttt 1083
Db 721 tcttcctcttctggaataattgctcactatagaaattgacttatgtcttactttt 780
QY 1084 tggtaactatggaataattgctcactatggttcttggtgagctggttagtgaacacacaa 1143
Db 781 tggtaactatggaataattgctcactatggttcttggtgagctggttagtgaacacacaa 840
QY 1144 aaacacaaagaagaattcccatatgaaaagaatacatcttcttcaatlaaagaat-- 1201
Db 841 aaacacaaagaagaattcccatatgaaaagaatacatcttcttcaatlaaagaatca 900
QY 1202 ----- 1201
Db 901 ggcgtgagtcagtagcatgatcttgctcaactgcagcctccctgcctccgggttca 960
QY 1202 -----caagcttcttcaagaagtc 1221
Db 961 agccatccctccacccctcagctcactcaagaagtatgttggaagcagagcttcttcaagaagtc 1020
QY 1222 agtgcgtggttaaccattttaaatacctcctgcacatttgggtatgattgacaact 1281
Db 1021 agtgcgtggttaaccattttaaatacctcctgcacatttgggtatgattgacaact 1080
QY 1282 tcttcaactgagacttttatacttattgacatatttgctcactatataagaagaat 1341
Db 1081 tcttcaactgagacttttatacttattgacatatttgctcactatataagaagaat 1140
QY 1342 cctaaggttcaatgttcaagaagaatggtttatcttcatattgaccttattagaccttg 1401
Db 1141 cctaaggttcaatgttcaagaagaatggtttatcttcatattgaccttattagaccttg 1200
QY 1402 gtatgtatgactcgtctgctgtaagcgtgctgacaatttaagggaacaaatgaaatttctc 1461

Dh 1201 gttatgatgatccctgctgctgaagcgtgctgacatlttaaggagcaaatgttttc 1260
QY 1462 aaccttgcttgctgaggaatttttaagcttataggaattgttgactgtgaattcc 1521
Dh 1261 aaccttaagtgctgcaggaatttttaagccttaaggaatgattggaacctgacatcc 1320
QY 1522 ggtagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1581
Dh 1321 ggtagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1380
QY 1582 aacacacacgtggagccttctgctgctgctgctgctgctgctgctgctgctgct 1641
Dh 1381 aacacacacgtggagccttctgctgctgctgctgctgctgctgctgctgctgct 1440
QY 1642 ttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1701
Dh 1441 ttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1500
QY 1702 tgggcccgtcatgtctaaagctggaacccctgataacactgttggagaatggaac 1761
Dh 1501 tgggcccgtcatgtctaaagctggaacccctgataacactgttggagaatggaac 1560
QY 1762 gttctatattgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1821
Dh 1561 gttctatattgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1620
QY 1822 aggtgaagtaacaaactggaacccctgataacactgttggagaatggaac 1881
Dh 1621 aggtgaagtaacaaactggaacccctgataacactgttggagaatggaac 1680
QY 1882 ataaataatccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1941
Dh 1681 ataaataatccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1740
QY 1942 tgtattgtgaagatctgctgctgctgctgctgctgctgctgctgctgctgctg 2001
Dh 1741 tgtattgtgaagatctgctgctgctgctgctgctgctgctgctgctgctgctg 1800
QY 2002 tataatcatgaataatcatcatgctgctgctgctgctgctgctgctgctgctgct 2061
Dh 1801 tataatcatgaataatcatcatgctgctgctgctgctgctgctgctgctgctgct 1860
QY 2062 taataatgaatgaatgctgctgctgctgctgctgctgctgctgctgctgctg 2121
Dh 1861 taataatgaatgaatgctgctgctgctgctgctgctgctgctgctgctgctg 1920
QY 2122 tgaagtgaagatctgctgctgctgctgctgctgctgctgctgctgctgctg 2181
Dh 1921 tgaagtgaagatctgctgctgctgctgctgctgctgctgctgctgctgctg 1980
QY 2182 accctcgtgataccttctgctgctgctgctgctgctgctgctgctgctgctgct 2241
Dh 1981 accctcgtgataccttctgctgctgctgctgctgctgctgctgctgctgctgct 2040
QY 2242 tgttgtgtgtcacttctcctcaataaattgtcagctcctcctgacacttagac 2301
Dh 2041 tgttgtgtgtcacttctcctcaataaattgtcagctcctcctgacacttagac 2100
QY 2302 tttagcatctctgtgagagcgcgcacacgtatatttgcgcgcgcgcgcgcgcgc 2361
Dh 2101 tttagcatctctgtgagagcgcgcacacgtatatttgcgcgcgcgcgcgcgcgc 2160
QY 2362 ggaatgtgcccagcagctgcccagcactccctccctgctgctgctgctgctgct 2421
Dh 2161 ggaatgtgcccagcagctgcccagcactccctccctgctgctgctgctgctgct 2220
QY 2422 cgtttatcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2481
Dh 2221 cgtttatcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2280
QY 2482 cctctgaggggctatcaatgtgtatgataaagccctgagtaggaagcagtagatca 2541
Dh 2281 cctctgaggggctatcaatgtgtatgataaagccctgagtaggaagcagtagatca 2340

QY 2542 ctgtcatggtctgtgatatacatccctcaaatctccctccacagcaggaatgtgctg 2601
Dh 2341 ctgtcatggtctgtgatatacatccctcaaatctccctccacagcaggaatgtgctg 2400
QY 2602 gcatgcaacctgcaaaaagaataatgcaagcgcgcgcgcgcgcgcgcgcgcgcgcgc 2661
Dh 2401 gcatgcaacctgcaaaaagaataatgcaagcgcgcgcgcgcgcgcgcgcgcgcgcgc 2460
QY 2662 cccagcacttggggggctgaggtggtggtggtggtggtggtggtggtggtggtggt 2721
Dh 2461 cccagcacttggggggctgaggtggtggtggtggtggtggtggtggtggtggtggt 2520
QY 2722 ggcagcagctggaaccccatctctctctctctctctctctctctctctctctctct 2781
Dh 2521 ggcagcagctggaaccccatctctctctctctctctctctctctctctctctctct 2580
QY 2782 cgggcgcctgtaatccagatactcaagagcctgagtaggaagaatcaactggaacctg 2841
Dh 2581 cgggcgcctgtaatccagatactcaagagcctgagtaggaagaatcaactggaacctg 2640
QY 2842 aggtggaagttgcaagtaacaaagatacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2901
Dh 2641 agatggaagttgcaagtaacaaagatacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2700
QY 2902 agactccaactc 2913
Dh 2701 agactccaactc 2712

RESULT 5
AAFS5900
ID AAF55900 standard; DNA; 2512 BP.
XX
AC AAF55900;
DT 18-APR-2001 (first entry)
XX
DE Human AST coding sequence.
XX
KW Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
KW anion and sugar transporter; anion-cation symporter;
KW Salla acid transporter; ss.
XX
OS Homo sapiens.
XX
PN EP1069184-A1.
PD 17-JAN-2001.
XX
PF 16-JUL-1999; 99EP-0202341.
PR 16-JUL-1999; 99EP-0202341.
XX
PA (ALKU) AKZO NOBEL NV.
XX
DR MPI; 2001-193090/20.
XX
P-PSDB; AAB66967.
PT New human transporter gene implicated in Salla disease and lysosomal
PT Salla acid transport, useful in assays for identifying new drugs, or
PT diagnosing Salla acid transport defects related to mutations in the
PT transporter gene -
XX
PS Claim 2; Page 12-13; 20pp; English.
XX
CC The present sequence is the coding sequence for human Anion and Sugar
CC Transporter (AST) protein. AST has significant homology with several
CC members of the anion-cation symporter (ACS) family of transporters. AST
CC is implicated in Salla disease, and is useful in screening assays for
CC identifying new drugs. Compounds identified via AST screening are useful
CC for preparing a pharmaceutical suitable as an activator or inhibitor of a
CC Salla acid transporter protein. The pharmaceutical may be used in Salla

|||||
Db 2046 aaagggatgagctgaaacagacctctgataaccttgcttaattaaactagatgataat 2105
QY 2221 tctaaagtaactgataaacaacctgtgtgttgcactctccatccataaaattgtagctctc 2280
Db 2106 tctaaagtaactgataaacaacctgtgtgttgcactctccatccataaaattgtagctctc 2165
QY 2281 tctgacacttagacacctaaacttagacatctctgttgagctgacatccactgtataatt 2340
Db 2166 tctgacacttagacacctaaacttagacatctctgttgagctgacatccactgtataatt 2225
QY 2341 cgcctggaactgactgagaggaggtgtgcccaagcagctgccaagcactccctctgg 2400
Db 2226 cgcctggaactgactgagaggaggtgtgcccaagcagctgccaagcactccctctgg 2285
QY 2401 cttaaggtcagagtgcccaacgcttatacagaagcagatcccaagcccaagcagagctc 2460
Db 2286 cttaaggtcagagtgcccaacgcttatacagaagcagatcccaagcccaagcagagctc 2345
QY 2461 gactctcggctggtgctcttctctctgaggggctatcatagtgtagataaagccctgagt 2520
Db 2346 gactctcggctggtgctcttctctctgaggggctatcatagtgtagataaagccctgagt 2405
QY 2521 aggcagaagcagtgagatccaactgtctatgtctgtatatactccaaacttccctccc 2580
Db 2406 aggcagaagcagtgagatccaactgtctatgtctgtatatactccaaacttccctccc 2465
QY 2581 aggcagaagcagatattgtctgagcagctgcaactgcaaaagaaaaattgc 2627
Db 2466 aggcagaagcagatattgtctgagcagctgcaactgcaaaagaaaaattgc 2512
RESULT 6
AAH79234
ID AAH79234 standard; cDNA; 2670 BP.
XX
AC AAH79234;
XX
DT 20-NOV-2001 (first entry)
XX
DE Human sodium dependent phosphate co-transfer protein 35 cDNA.
XX
KW Human; sodium dependent phosphate co-transfer protein 35;
KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN CNI298882-A.
XX
PD 13-JUN-2001.
XX
PE 06-DEC-1999; 99CN-0124217.
XX
PR 06-DEC-1999; 99CN-0124217.
XX
PA (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
XX WPI; 2001-503367/56.
DR P-PSDB; AAG65238.
XX
PT Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX
PS Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
XX
CC The present invention provides the protein and coding sequences of human
CC sodium dependent phosphate co-transfer protein 35. The sequences can be
CC used in the treatment of hypophosphaturia, hypercalcaemia,
CC hypophosphataemic rickets and nephritis. The present sequence is the
CC coding sequence of the invention.
XX

SQ Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;
Query Match 68.9%; Score 2018.6; DB 22; Length 2670;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2021; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 906 gagaagggtatatacttcacagcatgcatgcatgtgtctcttgggtccccccttg 965
Db 1 gagaagggtatatacttcacagcatgcatgcatgtgtctcttgggtccccccttg 60
QY 966 aaagaagcaaatctcttagcatcttgatgaagagacagcttggagacgaattcttc 1025
Db 61 aaagaagcaaatctcttagcatcttgatgaagagacagcttggagacgaattcttc 120
QY 1026 ttccctcttggaaataatttgctacatataatgaattgagcttattgtcttacttttg 1085
Db 121 ttccctcttggaaataatttgctacatataatgaattgagcttattgtcttacttttg 180
QY 1086 gtaactatggaataatttgcttctcttggatctgtttagttagtgacacacacaa 1145
Db 181 gtaactatggaataatttgcttctcttggatctgtttagttagtgacacacacaa 240
QY 1146 aacacaagagaattcccatatgaagaagaatacatcttccatcaataagaatacagc 1205
Db 241 aacacaagagaattcccatatgaagaagaatacatcttccatcaataagaatacagc 300
QY 1206 ttcttcaagaagaatcagtcggtgggtaccatctttaaataatccctggcacttgggcta 1265
Db 301 ttcttcaagaagaatcagtcggtgggtaccatctttaaataatccctggcacttgggcta 360
QY 1266 tctgtagttgacacatttcttcaactggaacttttatacttattgacataattggcta 1325
Db 361 tctgtagttgacacatttcttcaactggaacttttatacttattgacataattggcta 420
QY 1326 ctataatgaagagatcccttaagtccaatgttcaagaagaatggttttcttcaatgc 1385
Db 421 ctataatgaagagatcccttaagtccaatgttcaagaagaatggttttcttcaatgc 480
QY 1386 ctatttaaggctcttggtttattgatatgctgctgctgcaagcggcgaataattgaagg 1445
Db 481 ctatttaaggctcttggtttattgatatgctgctgctgcaagcggcgaataattgaagg 540
QY 1446 caaaatgaatttcttaacttattgtctgcaggaattttagacctatagaatgatagt 1505
Db 541 caaaatgaatttcttaacttattgtctgcaggaattttagacctatagaatgatagt 600
QY 1506 gacctgagatattccctgtgtagctgctgcttcaattgagctgtaattcttggccgttg 1565
Db 601 gacctgagatattccctgtgtagctgctgcttcaattgagctgtaattcttggccgttg 660
QY 1566 ctcttccatactataatacaacaacactggagagcttggctcttcttgatttagataaac 1625
Db 661 ctcttccatactataatacaacaacactggagagcttggctcttcttgatttagataaac 720
QY 1626 atctgatatgtctctctgtagctgtagtctccctgagcacaacaataacttggca 1685
Db 721 atctgatatgtctctctgtagctgtagtctccctgagcacaacaataacttggca 780
QY 1686 ctattccagaagaatggttggcccgctcaatgctaaagctgacccctgataaactgttg 1745
Db 781 ctattccagaagaatggttggcccgctcaatgctaaagctgacccctgataaactgttg 840
QY 1746 gagaatggcaaacggttctatattgctgctgctttaaagtgttggggcacttttc 1805
Db 841 gagaatggcaaacggttctatattgctgctgctttaaagtgttggggcacttttc 900
QY 1806 ttacactatcgcgcaaaagtgaagtaacaacaacggtgctccaatgatacaccagagaca 1865
Db 901 ttacactatcgcgcaaaagtgaagtaacaacaacggtgctccaatgatacaccagagaca 960
QY 1866 gacctggaaggaacaaataataatccctgctctatattgattttatcatgta 1925
|||||

Db 961 gacactgaaggaaccaataataatccctgctcctataatgtattttatatacatgta 1020
QY 1926 accctcaaaagtccctctctgtatgttgaagcattctatgctcttttttaattgta 1985
Db 1021 accctcaaaagtccctctctgtatgttgaagcattctatgctcttttttaattgta 1080
QY 1986 ttgattctttaaaggccataatcatatgaatatcactagctgtgcagagataataaagac 2045
Db 1081 ttgattctttaaaggccataatcatatgaatatcactagctgtgcagagataataaagac 1140
QY 2046 tctgtttaaataatgaataatgttaagctagagactctactacttaagttacaatccctgc 2105
Db 1141 tctgtttaaataatgaataatgttaagctagagactctactacttaagttacaatccctgc 1200
QY 2106 tctgtttaaataatgaataatgttaagctagagactctactacttaagttacaatccctgc 2165
Db 1201 tctgtttaaataatgaataatgttaagctagagactctactacttaagttacaatccctgc 1260
QY 2166 gaatgagctgaaacaacagactccctgataccttgcgttaataactagatgaataatcga 2225
Db 1261 gaatgagctgaaacaacagactccctgataccttgcgttaataactagatgaataatcga 1320
QY 2226 ggtactgataaacaacctgtgtgttgaactctccctcaataaataatgtcagctctctcga 2285
Db 1321 ggtactgataaacaacctgtgtgttgaactctccctcaataaataatgtcagctctctcga 1380
QY 2286 caactgagactcaaaactttaagatctctgtgagctgcgcatacctcgttaatttcgct 2345
Db 1381 caactgagactcaaaactttaagatctctgtgagctgcgcatacctcgttaatttcgct 1440
QY 2346 ggcgaactgagctgagggaggtgtgcaccaagcagctgcaccaagactccctccctgctca 2405
Db 1441 ggcgaactgagctgagggaggtgtgcaccaagcagctgcaccaagactccctccctgctca 1500
QY 2406 gggctcaagatgcccagcgtttatcagaagcagatcccaagcccagagcgaagtgtgcacc 2465
Db 1501 gggctcaagatgcccagcgtttatcagaagcagatcccaagcccagagcgaagtgtgcacc 1560
QY 2466 ttccgagctgtgctctcctcctctgagggcgtatcaatgtgtagataaaagccctgaagtgcga 2525
Db 1561 ttccgagctgtgctctcctcctctgagggcgtatcaatgtgtagataaaagccctgaagtgcga 1620
QY 2526 agagcagctgagatcccaactgtgtctgtatgatacactccctcaacttccctccagcac 2585
Db 1621 agagcagctgagatcccaactgtgtctgtatgatacactccctcaacttccctccagcac 1680
QY 2586 agagcagatctgtgctgctgcaacctgcagaaagaaaaatgcgagagcggcgagcaggt 2645
Db 1681 agagcagatctgtgctgctgcaacctgcagaaagaaaaatgcgagagcggcgagcaggt 1740
QY 2646 ggcctcaatgctttaaaccagcacttgggggctgaggtgtggcgagatagatgacag 2705
Db 1741 ggcctcaatgctttaaaccagcacttgggggctgaggtgtggcgagatagatgacag 1800
QY 2706 agttcgagaccagcgtgcagcagatgtgtgaaccccatctcactcaaaaaatacaaaaat 2765
Db 1801 agttcgagaccagcgtgcagcagatgtgtgaaccccatctcactcaaaaaatacaaaaat 1860
QY 2766 tagctggcgctgtgtgaagcgcgctgttaatcccgagatcagagagctgaggtagaga 2825
Db 1861 tagctggcgctgtgtgaagcgcgctgttaatcccgagatcagagagctgaggtagaga 1920
QY 2826 atcaacttgaacctgtggaaggttggaagtgtgagtggaacccaagatccagccctcagccag 2885
Db 1921 atcaacttgaacctgtggaaggttggaagtgtgagtggaacccaagatccagccctcagccag 1980
QY 2886 ccttgagcgagctgtgagcgagatcccaactcaaaaaaataaaaaa 2930
Db 1981 ccttgagcgagctgtgagcgagatcccaactcaaaaaaataaaaaa 2025

RESULT 7
AA250876

ID AA250876 standard; cDNA; 2844 BP.
AC AA250876;
XX 31-MAY-2000 (first entry)
DT 31-MAY-2000 (first entry)
DE Sheep GBS toxin receptor (SP55) cDNA.
KW Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritic; anti-psoriatic; ss.
XX Ovis sp.
XX
XX
FH Key Location/Qualifiers
FT CDS 84..1571 /tag= a
FT FT /product= "Sheep GBS toxin receptor"
FT FT
XX
XX WO200005375-A1.
PN 03-FEB-2000.
PD
XX
XX 22-JUL-1999; 99WO-US16676.
PR 22-JUL-1998; 98US-0093843.
XX
XX (UYVA-) UNIV VANDERBILT.
PI HELLERGVIST CG, Fu C;
XX
DR WPT, 2000-205377/18.
XX P-PSDB; AA145088.
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Claim 3; Page 83-86; 109pp; English.
XX
XX The present cDNA sequence encodes partial sheep GBS (group B beta
CC -haemolytic streptococci) toxin receptor (SP55). This sequence was cloned
CC using a primary culture of sheep lung endothelial cells. Expression
CC vectors comprising this cDNA can be transformed into host cells to
CC express GBS toxin receptor and its fragments. Detecting the receptor in
CC tissues is used to diagnose pathological vascularisation, e.g. for
CC detecting cancer metastases. GBS toxin receptors are useful for treating
CC conditions associated with pathological angiogenesis or
CC neovascularisation (specifically cancer, reperfusion injury, scarring
CC during wound healing, keloids, chronic inflammation (rheumatoid
CC arthritis or psoriasis) or neural injury), and to raise specific
CC antibodies used for treating early onset disease. Inhibitors of this
CC receptor are useful for treating pathological or hypoxia-induced
CC endothelial cell proliferation and migration.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

Query Match 53.6%; Score 1570.8; DB 21; Length 2844;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 1916; Conservative 0; Mismatches 387; Indels 33; Gaps 8;

QY 309 gccgcgggggttcgggtgtgtgcggcggtccctctctgcaggtgagtaaac 368
Db 5 gggcgcggggttcggcggttcgcgtgagctctcttcgcggagaggttgcgc 64
QY 369 ct-gctacgtagcgtatagatgtcttcggttcgagaccctgcggagagtgagga 426
Db 65 gtacgtccctgaagcatatgaatgcgcgttctcgacttagcccccagcagagcgga 124
QY 427 gtagagcaagcagcagcctctctacggcgccccaagcgccgaagcgcgtccagt 486

Db 125 ggagggctcggaacgcgcacacgcctccgcagcgccgcgcggcggaacccgttccagt 184
Qy 487 gtgtgtctgtcgtcgttacaaacttagcaatttggcctttttgtttcttcattgtga 546
Db 185 atgtctgtcgtcgttacaaacttagcaatttggcctttttgtttcttcattgtga 244
Qy 547 tgcattcgtgtgaatcgtgagtggttgcgttagtgagatgagatgaatacaactt 606
Db 245 ttcatcaggtgtgattcgtgagtggttgcgttagtgagatgaatacaacttgc 304
Qy 607 agaatgaataagaattccaagcggtgtccagagcattcgtctccctaaagtccca 666
Db 305 caaagaataatagaacgctcccaagagtgtagagagcattcgtctccctaaagtccca 364
Qy 667 taatcaaacggtgaagaagtaaccaattggatgtagcaaacccaagaagatcttcggttc 726
Db 365 caaccaaacgggtgaagaagtaaccaattggatgtagcaaacccaagaagatcttcggttc 424
Qy 727 cttttttatggtacatacatcacagatcctctgagagatagtgtgcagcaaatag 786
Db 425 tttttttatggtacatacatcacaaattcctgagagatagtgtgcagcaagaatg 484
Qy 787 gggaagaatcgtctgaagatttggatccttggacgtcgtcctcaacctgttcacttc 846
Db 485 gggaagcgtgtgctgaagattcggatccttgtctacagcattcctcaacctgttcacttc 544
Qy 847 catgtgcgcgaatttagaagatttggacccaactgtatcagaagcaatagaagaacag 906
Db 545 cctcgtcgcgaatttggatcggatcggagccctgtgtgcactcaaggcactaaagggctcag 604
Qy 907 agaggtgttacaattccacagacatgacgtcgtcctctgtggctccccccttga 966
Db 605 agaggtgttacaatccacagacatgacgtcgtcctctgtggctccccccttga 664
Qy 967 aagaagaagaactcttgaacttctgtatgtagagacacagcttggagcaatatttctc 1026
Db 665 aagaagaagaactcttgaacttctgtatgtagagagacacagcttggagcaatatttctc 724
Qy 1027 tccctctcgtgaataatttgcatactatgaattgagactatgctctcatttcttgg 1086
Db 725 tccctctcgtggaatttgcatactatgaattgagactatgctctcatttcttgg 784
Qy 1087 tactattgagaataatttgcatttcttcttggatcgtgtgaatgtagacacacacaaa 1146
Db 785 catgtgtggaatcattcgtgttattttagatcgtctgattagtagatcacacagaag 844
Qy 1147 acacaagaagaattcccatatagaagaagaatcattccttcacattagaagaatcagct 1206
Db 845 tcaacaagaacatcacccctcatagaagaagatatacttccatcacatataaatacagct 904
Qy 1207 tctctccagaagtcagtcggtggtgtaaccattttaaaatccctgcacttgggctat 1266
Db 905 cctctccagaagtcagtcggtggtgtaaccattttaaaatccctgcacttgggctat 964
Qy 1267 cgtatgtgcaacttctcttaacaacgtgacttttataactttatgaactatgtgctac 1326
Db 965 tgcgtgtgcaacttctcttaacaacgtgacttttataacttttgaactatgtgctac 1024
Qy 1327 ttatagaagaagatccttaagtggttcaagtccaagaagaagggttttattccttcattggcc 1386
Db 1025 ttcaagaagaagatccttaagtggttcaagtccaagaagaagggttttattccttcattggcc 1084
Qy 1387 ttatttgaactcgtgtgtatgtagtaccgtgtcgttgcagagcgtgcagcaaatlaagggc 1446
Db 1085 ttatttgaactcgtgtgtatgtagtaccgtgtcgttgcagagcgtgcagcaaatlaagggc 1144
Qy 1447 aaaaatggaacttccaacttatgtgttcgcagaatttttagcctatagaatgtagtgg 1506
Db 1145 aagaatggaacttccaacttatgtgttcgcagaatttttagcctatagaatgtagtgg 1204
Qy 1507 acctgagatcttcctgtgtagtctgctcattgctgtagtattattcttggccgttgc 1566
Db 1205 acctgagatcttcctgtgtgcgcaggaattatagcgtgtgattatccttggcgttgc 1264
Qy 1567 ttctcctaataatacaaacacactggaagcgttctgtcctctcgtatattagatcaacca 1626
Db 1265 attcctaataatacaaacacactggaagcgttctgtcctcgtatattagatcaacca 1324
Qy 1627 tctgagatgtctcctcgtatgtcgtgtatcctcctggtgcatcaacaatacattggccac 1686
Db 1325 tctgagatgtctcctcgtatgtcgtgtatcctcctggtgcatcaacaatacattggccac 1384
Qy 1687 tattccagaagatgttggggccgtcattgttaagaatgtcgaacccctgataactgttgg 1746
Db 1385 tattccagaagatgttggggccgtcattgttaagaatgtcgaacccctgataactgttgg 1444
Qy 1747 agaatgcaaacccgtgtctcataattcgtcgtgtatatagttttgtgtccatttctc 1806
Db 1445 agaatgcaaacccgtgtgtcgtcgtcgtcgtgtatcaatgatttttggccatttctc 1504
Qy 1807 tacaatttcgcgaaggtgaagtacaaaactgtggtcctcaatgataccaatgagacag 1866
Db 1505 cacactatcgcgaaggtgaagtacaaaactgtggtcctcaatgataccaatgagacag 1564
Qy 1867 acagtgaagaacccaataataatcctgcctcctattaagtattttattatcagttaa 1926
Db 1565 aaactgaagaacccaataataatcctgcctcctattaagtattttattatcagttaa 1624
Qy 1927 cctcaagtgaccttctcgtatgtgtgaagcattcctagt-----ccttttt 1972
Db 1625 cctcaagtgaccttctcgtatgtgtgaagcattcctagt-----ccttttt 1684
Qy 1973 aattgtactgtatgatttttaaggcctbaatacaatgaataatacattagttgcagaa 2032
Db 1685 tagaaaatgtgttagatttgaagcgtgtgataatcagtaagtaagtaagttgcacat 1744
Qy 2033 taataaatagaactgtgttaattatgaataataatgaactagcaggtcattttagt 2092
Db 1745 aagcaaaatagcatttcttataatataatcaacccgttgcgtgaacttacaacttcagg 1804
Qy 2093 tcaatacctcgtcgttagtcggtgcaacatgaagttaggaagttcgtgtattttag 2152
Db 1805 tcaatacctcgtcgttagtcggtgcaacatgaagttaggaagttcgtgtattttag 1863
Qy 2153 gccatacctaagggaatgagctgaacagacccctcgtataccttgcattatgaactag 2212
Db 1864 accatacctaagggaatgagctgaacatgaacccctcgtataccttgcattatgaactag 1923
Qy 2213 atgataatcctcaggtacgtataaacaacccgtgtgtgttcacattcctataaaa----- 2267
Db 1924 atataatcctcaggtacgtgtataaacaatcgttttgcataccttccataaaaattatt 1983
Qy 2268 -attgtcagctcctcgtacacttagacactcaaacctttagcatctctgtgagctgcat 2326
Db 1984 tgtcatcagaagaatccctgcagatgtaggtctcaaacctttagcctcctcaagagctggcag 2043
Qy 2327 ccaactgtataatctgcctcgtgcaactggaactggaaggtgtgtccagagcagctgccaa 2386
Db 2044 ccaactgtataatcagccctggaacttcaactctgaggggaagcagctgccagagcagctgccaa 2103
Qy 2387 caactcctcctcgtcctcagaggttcagaggtgccagaggtttatagaagcagcagctcaagc 2446
Db 2104 tgtccctcctcgtcctcagaggttcagaggtgccagagctt-----aggcagatccaaga 2156
Qy 2447 ccagagccaggtgtcgaactctcgtcgtggtccttccctctcgtgaggggtatcactgtgtag 2506
Db 2157 ccagggctcagcgcgaagcgttggagagatcttccct--ggggcgtttaatgtgtgg 2214
Qy 2507 ataaagccctgagtaggcaagagcagtgagatcactgctatggtctgtatatactccta 2566
Db 2215 atgaagccctgagtaggcaagagcagcagatccacagatcagtggtttccatgcacccct-- 2273
Qy 2567 aacttccctccagacagaggaatattgctgtgcatgcaacccctgcaaaaagaa 2622
Db 2274 tccctcctcctccagacacgtgagtagttcctgtgcagtgatacctgcgaagaagaa 2329

Db 1145 aagaatgaattttcaactctgtgtggtcgaagaatttttagccttaagagatgtg 1204
 QY 1507 accgcgagatctcctgttagctgtgcttcatgtgctgcatatcttggcgcgttg 1566
 Db 1205 accgcgagatctcctgtgtgtgcgcgaagattatagcgtgtatcttgcgcgttg 1264
 QY 1567 ttctctaactatatacaacaacactggaagccttgcctctctgtgatttagcatcaaca 1626
 Db 1265 attcctaaccatacaacaacccctggagagcttctgcctctctgtgatttagcatcaaca 1324
 QY 1627 tctgtgatatgtctcctctgtgtgtgtgtatcctcctgggcatcaataatctggcac 1686
 Db 1325 tctgtgacattgtctcctctgtgtgtgtgtatctcctgggcatcaataatctggcac 1384
 QY 1687 tatccagaagatgtgtggccgcgtcatgttaaaagtctgaacccctgaatacattgttg 1746
 Db 1385 tatcccggaatgtgtggcccatcattgccaagaagctcttaccctcgagaacacatttg 1444
 QY 1747 agaatggcaaacgctgtctatattgtcgtcgtatattgttttgggccatttctt 1806
 Db 1445 agaattgcacaactgttctctgcacgcgtcgtcatcatgtatttggcgcaatttctt 1504
 QY 1807 tacacttgcgcaagaagtgaagaaactggcctcaatgatacaacatgagacag 1866
 Db 1505 cacactatcgcgaaggtgaagtgcaaaactggccatcagtgatcaacaagaacacag 1564
 QY 1867 acactgaaggagaccaataaataactcgtcctctataatgtatttattatcatgtaa 1926
 Db 1565 aaactgaaggagaccaataaataactcgtcctctataatgtatttattatcatgtaa 1624
 QY 1927 cctcaagtgcctt-----ctgtatgtgtgaagcattcctatgt-----ctttttt 1972
 Db 1625 cctaaagctccttctgtatttattatgtgtgaagcaatcatalatacaagaataatgttac 1684
 QY 1973 aattgtactgtatagctatttttaggcctataatcgtataatcctactgttcgcaaa 2032
 Db 1685 tagaaaactgtgtctagatctgtgaagctgtgaacatgaaatgtcactgtgtgcata 1744
 QY 2033 taataaaatgaactgtgttataatataataatataatgaactgaagacttcaacttagt 2092
 Db 1745 aagcaaaattagctatttatttataatattataccgcttgcgtgaacttaacatcagg 1804
 QY 2093 tcaatatcctgcctcgtcagtcggaacatgtagagacagctcgtgtattttttag 2152
 Db 1805 tcaatatcctgcctcgtcagtcggaacatcgaacatagggagtgctctatla-tttaa 1863
 QY 2153 gccaataactgaactgaactgaacacacccctcgtatcccttggtttaataactag 2212
 Db 1864 accataactaaagatgagctgaataagacccctcctatcactcctgtcattaaagtg 1923
 QY 2213 atgataatctcagtgatgataaacaactgtgtgtgtcacttccataaana----- 2267
 Db 1924 ataataatctcagtgctgtttaaacaactcgtttttagtacacttccataaanaattatc 1983
 QY 2268 -atgttagctctctctgacacttaagacctcaaaacttaacatctcgtgtagctgccat 2326
 Db 1984 tgtatagaatccctgcgaatgtagctcacaactttagcctcctcacaagagctggcag 2043
 QY 2327 ccaactgataatctgcctgcgaactgagtgagtggtgtcccgagagcgcgaag 2386
 Db 2044 ccaactgataatcagcctgtgcgaactcactcgtgaggaagatgtcccgagagcgcga 2103
 QY 2387 caactcctcctgcctcagaggtcagaggtccagcgttattcagagcagcatccagc 2446
 Db 2104 tgtccctctcgtcctcagaggtcagaggtccagcgttattcagagcagcatccagc 2156
 QY 2447 ccagagcagctgtcagactcctgcgtgtccttccctcgtgagggcatcatatgtgtag 2506
 Db 2157 ccagagcagcgcgaagccttgcaggtatttctccctc-gggggtgtgaatgtgtg 2214
 QY 2507 ataaagccctgagtagcgaagagatgaatcactcgtcatgtgtcgtgataactccta 2566
 Db 2215 atgaagccctgagcgaacagagagcagcgtcatccacagatcgtgttccatgcacccctc- 2273

QY 2567 aacttccctccagcagaggaatattgctgtgcgtgacacttgcacaaagaa 2622
 Db 2274 tcccttccctccagcagcactgtgagtagctgtgcgtgacatgtacacttgcacaaagaa 2329

RESULT 9

AA158115
 ID AA158115 standard; cDNA; 1488 BP.

AA158115;

22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 318.

Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; chemokine; lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

OS Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

PE 26-DEC-2000; 2000MO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT;

PI WPI: 2001-442253/47.

DR P-PSDB; AAM38959.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

PS Claim 1; SEQ ID NO 318; 10078bp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cyostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localized neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, activity/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

Sequence 1488 BP; 368 A; 318 C; 323 G; 479 T; 0 other;

Query Match 50.7%; Score 1484.8; DB 22; Length 1488;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1486; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 386 atgaagctccggttcgagacctgcccggaaacgatacgagagagacgacgcgaag 445
 Db 1 atgaagctccggttcgagacctgcccggaaacgatacgagagagacgacgcgaag 60
 QY 446 cccctctacacggcgcccccacggcggaagccgcctcagctgagctgctgtcttcc 505
 Db 61 cccctctacacggcgcccccacggcggaagccgcctcagctgagctgctgtcttcc 120
 QY 506 aacttaagcaatttggcctttttgtttcttcatatgtatgcatatgagtgaaatcg 565
 Db 121 aacttaagcaatttggcctttttgtttcttcatatgtatgcatatgagtgaaatcg 180
 QY 566 agtgtgctgctagtgatgatagtatcaataacaactttagaagataatagaactcc 625
 Db 181 agtgtgctgctagtgatgatagtatcaataacaactttagaagataatagaactcc 240
 QY 626 aaggcggtgcacagagattctgtcccatataaagtcatcataatcaaacgggtaaag 685
 Db 241 aaggcggtgcacagagattctgtcccatataaagtcatcataatcaaacgggtaaag 300
 QY 686 taacaaatggatgcagaaacacgaagatgattctcggttccctttttatggctacac 745
 Db 301 taacaaatggatgcagaaacacgaagatgattctcggttccctttttatggctacac 360
 QY 746 atcaacagagattccctggagagatattgtccagcaaaaataggggggaaatgtcgtc 805
 Db 361 atcaacagagattccctggagagatattgtccagcaaaaataggggggaaatgtcgtc 420
 QY 806 ttggagatccctggacactgtgtccacccctggttaccacccatttctcagaattaga 865
 Db 421 ttggagatccctggacactgtgtccacccctggttaccacccatttctcagaattaga 480
 QY 866 gtgcgacacacatgatactacagagacactagaagagactaggaaggggtttacattcca 925
 Db 481 gtgcgacacacatgatactacagagacactagaagagactaggaaggggtttacattcca 540
 QY 926 gccatgcatgcatgtgtcttctcttgggtcccccctcttgaagaagaacacttcttagc 985
 Db 541 gccatgcatgcatgtgtcttctcttgggtcccccctcttgaagaagaacacttcttagc 600
 QY 986 atttgatagcaggagacagacttgggacagtaattcttcttcttcttcttcttcttctt 1045
 Db 601 atttgatagcaggagacagacttgggacagtaattcttcttcttcttcttcttcttctt 660
 QY 1046 tgcatacatggaattggaactatgtcttcttacttttttggtaactttggaattttgg 1105
 Db 661 tgcatacatggaattggaactatgtcttcttacttttttggtaacttttggtaattttgg 720
 QY 1106 ttctcttcttgatctggttagttagtgacacacacacacacacacacacacacacacac 1165
 Db 721 ttctcttcttgatctggttagttagtgacacacacacacacacacacacacacacacac 780
 QY 1166 tatgaagaagaataactttcttcatcataagaatcaagcttcttctcacaagaagtcatg 1225
 Db 781 tatgaagaagaataactttcttcatcataagaatcaagcttcttctcacaagaagtcatg 840
 QY 1226 ccgtggtgaccccaatttaaatccctgcacacttgggcatctgtagttgcacactttct 1285
 Db 841 ccgtggtgaccccaatttaaatccctgcacacttgggcatctgtagttgcacactttct 900
 QY 1286 taacaatggaacttttaactattatgacataatgtcctaactatataagaagaatccta 1345
 Db 901 taacaatggaacttttaactattatgacataatgtcctaactatataagaagaatccta 960
 QY 1346 aggttaaatggttcacagaagaatggtttatctatctatcttcttatttagctcttggta 1405
 Db 961 aggttaaatggttcacagaagaatggtttatctatcttcttatttagctcttggta 1020

QY 1406 tgaatgcctgtctggtcgaagctgtcgtgacaatttaaggcacaatggaatttttcaact 1465
 Db 1021 tgaatgcctgtctggtcgaagctgtcgtgacaatttaaggcacaatggaatttttcaact 1080
 QY 1466 ttatgtgttcgcagaattttttagccttataagaaatgatgttggacctgcagatcttcgta 1525
 Db 1081 ttatgtgttcgcagaattttttagccttataagaaatgatgttggacctgcagatcttcgta 1140
 QY 1526 gctgcgtgctcaattgctgctgtatattcttggccggttgcttcttcaactatataca 1585
 Db 1141 gctgcgtgctcaattgctgctgtatattcttggccggttgcttcttcaactatataca 1200
 QY 1586 acactggagagcttctgtcttcttcttggatattagatcaacacatctggtatctgtcctcg 1645
 Db 1201 acactggagagcttctgtcttcttcttggatattagatcaacacatctggtatctgtcctcg 1260
 QY 1646 tatgtcgtatcctcctgtggcatcaacaataatggtccacttccagaatgtgtgg 1705
 Db 1261 tatgtcgtatcctcctgtggcatcaacaataatggtccacttccagaatgtgtgg 1320
 QY 1706 cccgttcattgtctaaagctcgtgacccctgtatagaactgttggagaatggcaaccgtgttc 1765
 Db 1321 cccgttcattgtctaaagctcgtgacccctgtatagaactgttggagaatggcaaccgtgttc 1380
 QY 1766 tatatgtcgtcgtatataatgttttgggtccatttcttcaactatctgcgcaaaagt 1825
 Db 1381 tatatgtcgtcgtatataatgttttgggtccatttcttcaactatctgcgcaaaagt 1440
 QY 1826 gaagtcaaaactgggctcgaatgacacacatgacacagacactgta 1873
 Db 1441 gaagtcaaaactgggctcgaatgacacacatgacacagacactgta 1488

RESULT 10
 AA250880
 ID AA250880 standard; cDNA: 1485 BP.
 XX
 AC AA250880;
 XX
 DT 31-MAY-2000 (first entry)
 XX
 DE Human/Sheep consensus GBS toxin receptor cDNA sequence-1.
 KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
 KW pathological vascularisation; cancer metastases; angiogenesis;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
 OS Homo sapiens.
 OS Ovis sp.
 XX
 FH Key
 FT mat_peptide
 FT
 FT Location/Qualifiers
 FT 1..1485
 FT /tag= a
 FT /product= "Human/sheep consensus GBS toxin receptor
 FT protein sequence"
 FT /transl_except= (pos:4..6; aa:Xaa)
 FT /transl_except= (pos:7..9; aa:Xaa)
 FT /transl_except= (pos:16..18; aa:Xaa)
 FT /transl_except= (pos:22..24; aa:Xaa)
 FT /transl_except= (pos:28..30; aa:Xaa)
 FT /transl_except= (pos:31..33; aa:Xaa)
 FT /transl_except= (pos:34..36; aa:Xaa)
 FT /transl_except= (pos:46..48; aa:Xaa)
 FT /transl_except= (pos:49..51; aa:Xaa)
 FT /transl_except= (pos:58..60; aa:Xaa)
 FT /transl_except= (pos:61..63; aa:Xaa)
 FT /transl_except= (pos:64..66; aa:Xaa)
 FT /transl_except= (pos:67..69; aa:Xaa)
 FT /transl_except= (pos:70..72; aa:Xaa)
 FT /transl_except= (pos:73..75; aa:Xaa)

FT /transl_except- (pos:79..81; aa:Xaa)
FT /transl_except- (pos:85..87; aa:Xaa)
FT /transl_except- (pos:91..93; aa:Xaa)
FT /transl_except- (pos:100..102; aa:Xaa)
FT /transl_except- (pos:124..126; aa:Xaa)
FT /transl_except- (pos:130..132; aa:Xaa)
FT /transl_except- (pos:136..138; aa:Xaa)
FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..198; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..252; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)
FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)
FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)

FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)
FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1087..1089; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)
FT /transl_except- (pos:1348..1350; aa:Xaa)
FT /transl_except- (pos:1357..1359; aa:Xaa)
FT /transl_except- (pos:1372..1374; aa:Xaa)
FT /transl_except- (pos:1375..1377; aa:Xaa)
FT /transl_except- (pos:1381..1383; aa:Xaa)
FT /transl_except- (pos:1384..1386; aa:Xaa)
FT /transl_except- (pos:1396..1398; aa:Xaa)
FT /transl_except- (pos:1402..1404; aa:Xaa)
FT /transl_except- (pos:1420..1422; aa:Xaa)
FT /transl_except- (pos:1444..1446; aa:Xaa)
FT /transl_except- (pos:1456..1458; aa:Xaa)
FT /transl_except- (pos:1459..1461; aa:Xaa)
FT /transl_except- (pos:1462..1464; aa:Xaa)
FT /transl_except- (pos:1471..1473; aa:Xaa)
FT /transl_except- (pos:1483..1485; aa:Xaa)
FT /note="Xaa = unknown"

XX WO200005375-A1.
 XX
 PN 03 - FEB- 2000.
 PD
 XX
 XX 22 - JUL- 1999; 99WO-US16676.
 PF
 XX 22 - JUL- 1998; 98US-0093843.
 PR
 XX
 XX (UYVA-) UNIV VANDERBILT.
 PA
 XX
 XX
 PI Høllergvist CG, Fu C;
 XX
 XX WPI, 2000-205377/18.
 DR P-PSDB; AAY45090.
 DR
 XX

Query Match	48.28;	Score 1413;	DB 21,	Length 1485;
Best Local Similarity	87.98;	Pred. No. 0;		
Matches 1305; Conservative	180;	Mismatches	0;	Gaps 0;

OY	386	atgaggtctccggttcgaagacatcgcccggaaagaatgtagaagacaagacgcgaag	445
Db	1	atgargtgcgcggttttyrargaatrrtgccsgaracgayaaggagrgcwgvgccgcac	60
OY	446	cctcttaccggcgcccccgccgagccgaagccgctccagtgtgcgtctgtcttac	505
Db	61	ckctkctrcmgsgscgcccccctggcgcsagaascgcgtcccastrrgctgtctgtcttac	120
OY	506	aacttaagcaatttgcgcttttgytttctccaatttgatgatcatagctgtgaactg	565
Db	121	aacytaagawttttgacctttttgtttctctcrststatkcatragkygfaatcg	180
OY	566	agcttgcgttaagcgatgatgatataaatcaaacctttaagaataataaacttc	625
Db	181	agyltgcrgytagtgyagatbygrtatlttcaaayacaactkymraadgaataaagaacttc	240
OY	626	aagcgctgccagagcattcgcgtccccaataaadtatcataataacaaggttaagaag	685
Db	241	wasgmrgtscagagcattctgcctcccataaaagtlctwtaayaacaaaggttaaraag	300
OY	686	taccatagygatugcagaacaactcaagatalgatlctgcgtlctcttttlaibgctacatc	745
Db	301	taccrrtgagatgagaacaactcaagatlgatcttcgwcycytttttyatbtgctaactc	360
OY	746	atcacacaagatcctgtagagatatbgttgcacgacaaataaggggggaaalqtctgtatga	805
Db	361	atcacacaaratctctggagatatagttgcacgaeakraakwggggggaarmgytgcataga	420
OY	806	tttggagatccttgacatgctgcgtccctacaccttcaactcccatigtctgaagatttaga	865
Db	421	tttyggagatcytttgyaacgcgctrrctytacacctgttcaactcccmtygcugagattcmga	480
OY	866	gttggaccaactcatgttgaactoaagacactaagaaagaaactlagagaaggtgttaacttcca	925
Db	481	gttygasacomtyrtiyactoaagtagacactagaagagrcrtlagagaggggtgyacabwtcca	540
OY	926	gcacatgcacgcacgtggtcttcttgygctcccccctctyaagaagacaaacttcttagc	985
Db	541	gcacatgcacgcacgtggtcttcttgygctcccccctcttlaaagaagacaartctckagy	600
OY	986	attccgatgcagagagacagcctgtggagaagtaattctctctctcttccttgtgaataat	104
Db	601	attccratgcagagagacacaacttggagacaglatcttctctccctcttcygarfaatt	660
OY	1046	tgctactatataatitggaactatgctcttactcttttggtaactatitggaactatttg	110
Db	661	tgctacataatagatttgacttaagctctctatyttytggyyrttttggaalmwytyg	720
OY	1106	ttctcttttgyagatcyggttagttagtgcacacacacaaacaaacagaagaattccocat	116
Db	721	tttttttttrtgagctbsttagtttagttagyavacacacaaamwccacagaagaaalywycm	780

QY	1166	tatgaagaaggaatcacatctcttcaatcaatgaagaatcagcttcttccacagaagtcagtg	1225
Db	781	tatgaagaagatgatatctcttccatcatctaaaraaacagctctctccacagaagtcagtg	840
QY	1226	cctgctggatccacattttaaaatccctgcacatttgggtatcgtagttgcacattttct	1285
Db	841	cctgtrtactcagcatktyrtaaatcmctgcacatttgggtatcgtagttgmgatgcaylttct	900
QY	1286	tacacatgcagcttttatactattatgcacattatgcctacttataatgaagaagatccta	1345
Db	901	tacacatgcagcttttataactttrtgcgmatatgtcctacttayaagaagarrtcccta	960
QY	1346	aggttcaaatgttcaagaagaatcgggttttatactatcatcctacttataatgagtcctgttta	1405
Db	961	aggttcaaatrrttcaagaagaatcgggttttatactatcctcaakscctattatgylstgttta	1020
QY	1406	tgtatgatccctgtctcgttccaagctgcctgcacaatttaaggcacaatggaaattttcaact	1465
Db	1021	tgtatgatccctgtctcgttccaagctcgtcacaatttaaggcacaatggaaattttcaact	1080
QY	1466	ttagtatttttcgaagaattttttagacctttaagaagaatttggaccctgcctactcctgtta	1525
Db	1081	yrtrtktgtctcgagaarrtttttagaccttaagaagaatttggaaactcgtcrrtatccctggtw	1140
QY	1526	gctctgtgccttcatttgcctgcgtgatattcttctggccgttgccttctccactacataca	1585
Db	1141	gcggcgwgmtyatwtgscgtctgatattcttctgtgcgtytgcwtctccactaayataca	1200
QY	1586	acactcgggaagcctttgtcctctcttcgtgatttgacatcaacatctcgtatatgtctcctcg	1645
Db	1201	acmctcgggaagcctttgtcctctcttcgtgatttgacatcaacatctcgtayatgtcctcctcg	1260
QY	1646	tatgtctgtatccctccctgggcatcaacaatatatttgcacattcccgaggaatggtcggg	1705
Db	1261	tatgtctgtatcctccctgggcatcaacaatcmittgcacattcccgaggaatggtcggg	1320
QY	1706	ccgcgcattgtctaaagctcgtaccctctatatacactcgttggagaatggcacaacgtgttc	1765
Db	1321	ccgcrtactgtayaraagctcctkaccctctgakaacacrttrtggagaatggcacaacgtgttc	1380
QY	1766	tatatctgtcgtctcattaaatggtttttggtgcacatttcttcaactattgcaccaagtg	1825
Db	1381	trtatgtcgtcgtcgtcatyaaatggtttttggtgcacatttcttcaactattgcaccaagtg	1440
QY	1826	gaagracaaaaactggcctcgaatgatatacatatggacacagaac	1870
Db	1441	gaagracaaaaactggcctcgaatgatatacatcawgacacagamac	1485
RESULT 11			
AAH99626/C			
ID	AAH99626 standard; cDNA: 1975 BP.		
XX	AAH99626;		
DT	16-OCR-2001 (first entry)		
XX	Human protein encoding cDNA sequence SEQ ID NO:461.		
DE	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;		
KM	antituberculosis; antitubercular; antitubercular; immunosuppressive;		
KM	antibacterial; endocrine; cardiac; central nervous system; virulence;		
KM	anti-HIV; fungicide; antitumor; cardiovascular; antineoplastic; anaemia;		
KM	antitubercular; haemostatic; vulnery; antituberc; osteopathic; eczema;		
KM	dermatological; antiallergic; antitubercular; antidiabetic; cytostatic;		
KM	neuroprotective; antipressant; nootropic; antiparkinsonian; infection		
KM	immunostimulant; gene therapy; antitense therapy; vaccine; inflammation		
KM	antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;		
KM	cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;		
KM	genetic disease; haematopoietic disorder; platelet disorder; asthma;		
KM	thrombocytopenia; osteoporosis; severe combined immunodeficiency;		
KM	allergic rhinitis; diabetes; multiple sclerosis; depression;		
KM	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;		

KW neurological disorder; ss.
XX Homo sapiens.
OS WO200153455-A2.
XX 26-JUL-2001.
XX 22-DEC-2000; 2000WO-US35017.
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
XX Tang YF, Liu C, Drmanac RT;
PI WPI: 2001-457603/49.
DR P-PSDB; AAM25685.
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 1; Page 537; 1217pp; English.
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virocidic; anti-HIV; fungicidal; antitubercular;
CC cardiovascular; antidiabetic; antiallergic; haemostatic; vulnery;
CC antitubercular; cytosolic; dermatological; antiallergic; antisthmatic;
CC antidiabetic; cytosolic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX Sequence 1975 BP; 535 A; 437 C; 434 G; 569 T; 0 other:
SQ

Query Match 48.0%; Score 1405.6; DB 22; Length 1975;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1479 gaattttagcctatagaatgagatgagcctgacagatctctgtagctgctgctta 1538
DB 1975 GAATTTTATGCTTATAGGAATGATTTGACCTGACATATTCCTGATGCTGCTTCA 1916
QY 1539 ttgctgtgattatctcttgccgttgcttccctaactatcaacaacacgtggagct 1598
DB 1915 TTGGCTGTGATTTATCTTGGCCGCTTCTTCTTAATATCAACAAACACCTGGAGCT 1856
QY 1599 ttgctctcttgatttagcatcaacacatgatatgctctctctgtagctgctatcc 1658
DB 1855 TTTCCTCTTCTGATTTACATCAACCATCTGATATTCCTCTCTGATGCTGATTC 1796
QY 1659 tccctggacacacaataacattgccaactatccagaatggttgagccgcatatgcta 1718
DB 1795 TCCCTGGGCAATCAACAATATTCATTTGCCACATATTCAGAGATGCTGGCCGCTCATTTGCTA 1736
QY 1719 aaagtcgacccctgataacactgttgagaatggaacacgctgtctatctgctgctg 1778
|||||

DB 1735 AAGCTGACCCCTGATTAACACTGTTGGAGAAATGGCAAAACCGTTCATATTCGCTG 1676
QY ctatgaatgtttttgtagcatttctta--cactatccgcaaaagtgagttacaata 1836
DB 1675 CTATTAATGTTTGGTGGCCATTTTCTTAACACTATTTCCGCAAAAGGTGAAGTACAAAA 1616
QY ctggagctcaatgatacaacatgagacacagacactgaaggaaccaaataaactctgcc 1896
DB 1615 CTGGGCTCAATGATGACCATGAGACAGACACTGAAGAACCAATTAATATCTCGCC 1556
QY 1897 tctattaaagtatttattatcatgataccctcaagtgccctctgtagttagtaaga 1956
DB 1555 TCTATTAATGATTTTATTTATTCATGATACCTCAAGTGCCCTCTGATTTGGTAAGCA 1496
QY tctatgctcttttcttaactgtagttagttagtatttaagggcctataatcaatgaata 2016
DB 1495 TTCTATGCTCTTTTAAATGTAAGTGTATGATTTTAAAGCCATATATCATGTAATA 1436
QY 2017 tcaatgattgccaagaataaataaagaactgtgttaattatgaataatglaagctag 2076
DB 1435 TCACATGTTGGCAGATTAATAAATGAACGTGTATTAATTATGAATAATATGTAACCTAG 1376
QY 2077 gaattctacttagtgatcaatacctgctgctgtagtcgaggaatgaaatggaagat 2136
DB 1375 GACTTCTACTTTAAGTTCAACATACCTGCTGTAGTCGGGCAACATGAAGTAGACAGTT 1316
QY 2137 ctgtgatttttttagggccaactaagaaggaatgagctgaacaacacccctgatacct 2196
DB 1315 CTGTATATTTTATGAGGCCATATCAAGGGAATGAGCTGAACAGCTCTGATACCTT 1256
QY 2197 tgcctaatlaactagatgaataatctcagtagttagtaataaacactgtgtgttact 2256
DB 1255 TGCTTATTAATTAACATGATATATTCAGTACTGATTAACACCTGTTGTTGCTACCTT 1196
QY 2257 tctctataaattgtcagctctctctgacacttagaactcaacttagactctctg 2316
DB 1195 TCCCTCTAATAAATTTGACACTCTCTCTGACACTTAACCTCAAACTTGTGACTCTG 1136
QY 2317 gaggctccatccactgataatctgcccgtgcaacgtgagctgaaggaatgtagccagagc 2376
DB 1135 GAGCTGCCATCCACTTATATTTGCGCTGGCAACGTGAGCTGAGGAGATGGCCAGGC 1076
QY 2377 agctgcaagcactccctccctgctgcttagaggtcagaagtgcacagctttatcaagagca 2436
DB 1075 AGCTGCCAAGCACTCCCTCCCTGCTTCAAGGCTCAAGAGTCCAGAGCTTATCAGAGCA 1016
QY 2437 gcatccaaagccagagcaggtgtgactcttgctgctgctgcttctcttgaggagctat 2496
DB 1015 GCATCCAAAGCCAGAGCCAGTGCACCTCTTGCGCTGCTGCTTCTCTGAGGGGCTAT 956
QY 2497 caatgtatgataaagccctgtagtaagcaagaagctgagatccactgctatgcttga 2556
DB 955 CAATGTGTATGATTAAGCCCTGAGTAGGCAAGAGCACTGAGATCCATGCTATGCTTTTA 896
QY 2557 taacatccaaacttccctccctccagacagaggaatattgctgcatgcaactgcaaa 2616
DB 895 TACATCTCAAACTTCCCTCCACACAGAGAAATATTGCTGGCATGCAACCTGCAA 836
QY 2617 aagaataaattgcaagcggcgagcagcagctgctatgctgctgaatccagacacttgagg 2676
DB 835 AAGAAAAATGCAAAAGGGCGGGGCAAGGCTGCTGTAATGCCAGCACTTGGGG 776
QY 2677 ggcctagatgagcgaatactgagatcagaagatctgaagaccacactgagcagaatgtaga 2736
DB 775 GGCCTAGGTGGCGAATCTAGAGATCAGAGATTCCAGACCACTGGCCAGCATGGTGA 716
QY 2737 accccatctactaataaataacaaaatlaactgagcgttgtagcagcgccctgtaac 2796
DB 715 ACCCATCTCTCTATAAATTAACAAAAATTTACCTGGCGCTGTGAGCGGCGCTGTAAATC 656
QY 2797 ccagataactcaggaagctgaggttagagagaatcaacttgaacctgaggaagttagaagctgag 2856
DB 655 CCAGATACTCAGAGAGCTGAGGTAGAGATCACTTGAACCTGAGGAGATGAGATGAGTTGAG 596
|||||

FT	/transl_except=	(pos:1012..1014; aa:Xaa)
FT	/transl_except=	(pos:1033..1035; aa:Xaa)
FT	/transl_except=	(pos:1063..1065; aa:Xaa)
FT	/transl_except=	(pos:1081..1083; aa:Xaa)
FT	/transl_except=	(pos:1084..1086; aa:Xaa)
FT	/transl_except=	(pos:1090..1092; aa:Xaa)
FT	/transl_except=	(pos:1096..1098; aa:Xaa)
FT	/transl_except=	(pos:1111..1113; aa:Xaa)
FT	/transl_except=	(pos:1126..1128; aa:Xaa)
FT	/transl_except=	(pos:1129..1131; aa:Xaa)
FT	/transl_except=	(pos:1138..1140; aa:Xaa)
FT	/transl_except=	(pos:1141..1143; aa:Xaa)
FT	/transl_except=	(pos:1144..1146; aa:Xaa)
FT	/transl_except=	(pos:1147..1149; aa:Xaa)
FT	/transl_except=	(pos:1150..1152; aa:Xaa)
FT	/transl_except=	(pos:1153..1155; aa:Xaa)
FT	/transl_except=	(pos:1168..1170; aa:Xaa)
FT	/transl_except=	(pos:1174..1176; aa:Xaa)
FT	/transl_except=	(pos:1180..1182; aa:Xaa)
FT	/transl_except=	(pos:1189..1191; aa:Xaa)
FT	/transl_except=	(pos:1201..1203; aa:Xaa)
FT	/transl_except=	(pos:1246..1248; aa:Xaa)
FT	/transl_except=	(pos:1270..1272; aa:Xaa)
FT	/transl_except=	(pos:1291..1293; aa:Xaa)
FT	/transl_except=	(pos:1306..1308; aa:Xaa)
FT	/transl_except=	(pos:1315..1317; aa:Xaa)
FT	/transl_except=	(pos:1324..1326; aa:Xaa)
FT	/transl_except=	(pos:1330..1332; aa:Xaa)
FT	/transl_except=	(pos:1333..1335; aa:Xaa)
FT	/transl_except=	(pos:1339..1341; aa:Xaa)
FT	/transl_except=	(pos:1348..1350; aa:Xaa)
FT	/transl_except=	(pos:1357..1359; aa:Xaa)
FT	/transl_except=	(pos:1372..1374; aa:Xaa)
FT	/transl_except=	(pos:1375..1377; aa:Xaa)
FT	/transl_except=	(pos:1381..1383; aa:Xaa)
FT	/transl_except=	(pos:1384..1386; aa:Xaa)
FT	/transl_except=	(pos:1396..1398; aa:Xaa)
FT	/transl_except=	(pos:1402..1404; aa:Xaa)
FT	/transl_except=	(pos:1420..1422; aa:Xaa)
FT	/transl_except=	(pos:1444..1446; aa:Xaa)
FT	/transl_except=	(pos:1456..1458; aa:Xaa)
FT	/transl_except=	(pos:1459..1461; aa:Xaa)
FT	/transl_except=	(pos:1462..1464; aa:Xaa)
FT	/transl_except=	(pos:1471..1473; aa:Xaa)
FT	/transl_except=	(pos:1483..1485; aa:Xaa)
FT	/note="Xaa = unknown"	

Query Match	44.5%	Score 1305;	DB 21;	Length 1485;
Best Local Similarity	87.9%	Pred. No. 1,1e-310;		
Matches 1305;	Conservative	0;	Mismatches 180;	Indels 0;
				Gaps 0;
QY 386 atgaagctccagttcgcagactgcgcgcgaacatgtgcgaagagacagaccacgacg				
Dd 1 atgaagctccagttcgcagactgcgcgcgaacatgtgcgaagagacagaccacgacg				
QY 446 cctcttaccgcgcgcgcccaacgcgcgaacgcgtccagttgtctgtctcgtcttac				

[illegible]


```
Qy 1060 ttgagactatgtctctactttttgttactatgtgaatatltgtgttcttcttgat 1119
      |||
Db 793 ttgactatgtctctactttttgttactatgtgaatatltgtgttcttcttgat 852
Qy 1120 ctggttctgttaagtgacacacacacacacagaagaatttcccatatgtaaaaagata 1179
      |||
Db 853 ctggttctgttaagtgacacacacacacacagaagaatttcccatatgtaaaaagata 912
Qy 1180 catcttcatcatat 1196
      |||
Db 913 catcttcatcatat 929

RESULT 14
AAK93901
ID AAK93901 standard; cDNA; 853 BP.
AC AAK93901;
AT 06-NOV-2001 (first entry)
DE Human cDNA clone representative sequence, SEQ ID NO: 2361.
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX Homo sapiens.
OS EP1130094-A2.
PN 05-SEP-2001.
PD 07-JUL-2000; 2000EP-0114089.
PE 08-JUL-1999; 99JP-0194486.
PF 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000UP-0183765.
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PS Example 11; SEQ ID NO 2361; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 853 BP; 198 A; 200 C; 213 G; 239 T; 3 other;
```

```
Query Match 28.3%; Score 827.8; DB 22; Length 853;
Best Local Similarity 98.7%; Pred. No. 1.7e-193;
Matches 843; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
```

```
Qy 304 agctgcgccggggtctggtctgttcggcgccgctccctctctctgcaggtgagag 363
      |||
Db 1 agctgcgccggggtctggtctgttcggcgccgctccctctctctgcaggtgagag 60
```

```
Qy 364 tacacctgtcacgtaagcgctcatgaaggtctccggttccgaacctggcccggaacgtg 423
      |||
Db 61 tacacctgtcacgtaagcgctcatgaaggtctccggttccgaacctggcccggaacgtg 120
Qy 424 cgaaggaagcagcgaccgcgcctctctacccggcgccccacggcgcaagcgcctc 483
      |||
Db 121 cgaaggaagcagcgaccgcgcctctctacccggcgccccacggcgcaagcgcctc 180
Qy 484 agtgtgtgtctgtctgtctgtacactagaatttggcctttttgtttcttcattgt 543
      |||
Db 181 agtgtgtgtctgtctgtctgtacactagaatttggcctttttgtttcttcattgt 240
Qy 544 gtatgcattcgttgaatctgagtggttgcttagtgaatgtgaatcacaataaac 603
      |||
Db 241 gtatgcattcgttgaatctgagtggttgcttagtgaatgtgaatcacaataaac 300
Qy 604 tttagaagataatagaacttcccaaggcgtgtccagagacattctgtcccaataagltca 663
      |||
Db 301 tttagaagataatagaacttcccaaggcgtgtgtccagagacattctgtcccaataagltca 360
Qy 664 tcataatcaaacgggtgaagatgtaaccaatgtggatgcaagaactccaaggatgtctcg 723
      |||
Db 361 tcataatcaaacgggtgaagatgtaaccaatgtggatgcaagaactccaaggatgtctcg 420
Qy 724 ttcccttttttactgtgtacatacacaagatctcctgtgaggtatgtgtccagcaaat 783
      |||
Db 421 ttcccttttttactgtgtacatacacaagatctcctgtgaggtatgtgtccagcaaat 480
Qy 784 aggggggaaaaatgctgtgaagatttggaatcctgtgcaactgtgtcccaacctgtcac 843
      |||
Db 481 aggggggaaaaatgctgtgaagatttggaatcctgtgcaactgtgtcccaacctgtcac 540
Qy 844 tcccatgtgcgagatttgaagatttgacaccatgtgtactgaaggacataaaggact 903
      |||
Db 541 tcccatgtgcgagatttgaagatttgacaccatgtgtactgaaggacataaaggact 600
Qy 904 aggaagaggtgtacatttcccaagcaltgcacatgtgtcttcttctgtggtccctct 963
      |||
Db 601 aggaagaggtgtacatttcccaagcaltgcacatgtgtcttcttctgtggtccctct 660
Qy 964 tgaagaagcaaaccttcttagcatctcgatcagagagcagacgttggagcaatcttc 1023
      |||
Db 661 tgaagaagcaaaccttcttagcatctcgatcagagagcagacgttggagcaatcttc 720
Qy 1024 tcttcctcttctggaataatttgcactataagaattgactatgtcttcttcttct 1083
      |||
Db 721 tcttcctcttctggaataatttgcactataagaattgactatgtcttcttcttctt 780
Qy 1084 tggactatgtgaatatltgtgttcttctgttgatcgtgttagttagtacacacaca 1143
      |||
Db 781 tgnactatgtgaaaaatttgggttcttcttnggaltctngtag-tagtagcacacacaca 859
Qy 1144 aaaaacaagaagaa 1157
      |||
Db 840 aaaaaccacagaagaa 853
```

```
RESULT 15
AAK92364
ID AAK92364 standard; cDNA; 838 BP.
XX
XX AAK92364;
```

```
AT 06-NOV-2001 (first entry)
```

```
DE Human cDNA 5'-end sequence, SEQ ID NO: 824.
```

```
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
```

```
OS Homo sapiens.
XX EP1130094-A2.
```



```
XX 05-SEP-2001.
PD
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PI use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 824; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 838 BP; 189 A; 196 C; 210 G; 237 T; 6 other;

Query Match      26.5%; Score 775.6; DB 22; Length 838;
Best Local Similarity 97.8%; Pred. No. 1.2e-180;
Matches 795; Conservative 0; Mismatches 17; Indels 1; Gaps 1;
```

```
QY 784 aggggggaaatgctgtaggaattgggaacctggcactgctgtccctcaacctgttcac 843
   |||
DB 481 aggggggaaatgctgtaggaattgggaacctggcactgctgtccctcaacctgttcac 540
QY 844 tcccatgtcgcagatlttagaggttgagaccactcaattgtactcagaagcactgaaggact 903
   |||
DB 541 tcccatgtcgcagatlttagaggttgagaccactcaattgtactcagaagcactgaaggact 600
QY 904 aggaaggggtgttacattccagcatgcatgcatgctgtctcttgggctcccccctt 963
   |||
DB 601 angagaggggtgttacattccagcatgcatgcatgctgtctcttgggctcccccctt 660
QY 964 tgaagaagcaactctctagcatlctgtatgcagagacagacttgagacagtaatttc 1023
   |||
DB 661 tgaagaagcaactctctagcatlctgtatgcagagacagacttgagacagtaatttc 720
QY 1024 tcttcctcttctgaaataattgctactataatgaatgactatgtcttactttt 1083
   |||
DB 721 tcttctcttctgaaataa-ttgcactataatgaatgactatgtcttactttt 779
QY 1084 tggactatggaatatatttggttctcttctgtg 1116
   |||
DB 780 tggactatggaatatatttggttctcttctgtg 812
```

Search completed: July 15, 2002, 13:52:14
Job time: 14874 sec

• • •

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 09:40:35 ; Search time 6938.09 Seconds
(without alignments)
8837.411 Million cell updates/sec

Title: US-09-776-865-1

Perfect score: 2930

Sequence: 1 gtcggtcgagccctcccc.....ctcaaaaaaaaaaaaaaaaaa 2930

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	2930	100.0	2930	6	AX207624	AX207624 Sequence
2	2930	100.0	2930	9	AF244577	AF244577 Homo sapi
3	2619.2	89.4	3329	9	BC020961	BC020961 Homo sapi
4	2359.4	80.5	2512	6	AX138494	AX138494 Sequence
5	2359.4	80.5	2512	9	HS387747	AJ387747 Homo sapi
6	1570.8	53.6	2844	4	AF244578	AF244578 Ovis arie
7	1570.8	53.6	2844	6	AX207626	AX207626 Sequence
8	1565.8	53.4	1587	9	AK025880	AK025880 Homo sapi
9	1336.4	45.6	2006	9	AK026921	AK026921 Homo sapi
10	1193.6	40.7	5664.1	2	AL138833	AL138833 Homo sapi
11	1193.6	40.7	113202	9	HSJ397H23	AL121972 Human DNA
12	1192	40.7	149597	2	AC034271	AC034271 Homo sapi
13	1192	40.7	157749	2	AC025535	AC025535 Homo sapi
14	400	13.7	163577	9	AL590428	AL590428 Human DNA
15	396.8	13.5	149597	2	AC034271	AC034271 Homo sapi
16	375	12.8	375	6	AX118967	AX118967 Sequence
17	285.2	9.7	179129	2	AC097023	AC097023 Rattus no
18	254.8	8.7	174852	2	AC010195	AC010195 Homo sapi
19	254.4	8.7	160696	9	AC090937	AC090937 Homo sapi
20	254.4	8.7	177632	2	AC024098	AC024098 Homo sapi
21	254.4	8.7	181321	9	AC093179	AC093179 Homo sapi
22	254.4	8.7	193700	2	AC024545	AC024545 Homo sapi
23	253.2	8.6	166625	2	AC090676	AC090676 Homo sapi
24	253.2	8.6	166685	2	AC055123	AC055123 Homo sapi
25	253.2	8.6	189968	2	AC013419	AC013419 Homo sapi
26	253	8.6	7784	9	AC093035	AC093035 Homo sapi
27	252.8	8.6	88281	9	AC095067	AC095067 Homo sapi
28	252.8	8.6	160434	2	AC018998	AC018998 Homo sapi
29	252.8	8.6	168229	2	AC104696	AC104696 Homo sapi
30	252.6	8.6	156420	9	AC007510	AC007510 Homo sapi
31	252.6	8.6	178328	9	AC092547	AC092547 Homo sapi
32	252	8.6	158608	2	AC026834	AC026834 Homo sapi
33	252	8.6	179531	9	AC003080	AC003080 Human BAC
34	252	8.6	220399	2	AC015918	AC015918 Homo sapi
35	251.8	8.6	189883	9	AC009075	AC009075 Homo sapi
36	251.4	8.6	3946	9	AB032435	AB032435 Homo sapi
37	251.4	8.6	118396	9	AC073138	AC073138 Homo sapi
38	251.4	8.6	155929	2	AC025574	AC025574 Homo sapi
39	251.4	8.6	174305	9	AL450327	AL450327 Human DNA
40	251	8.6	37773	9	AF027207	AF027207 Homo sapi
41	251	8.6	180063	2	AC074281	AC074281 Homo sapi
42	251	8.6	186006	2	AC073989	AC073989 Homo sapi
43	251	8.6	340000	9	HS21C067	AL163267 Homo sapi
44	251	8.6	348050	9	HS229042	AJ229042 Homo sapi
45	250.8	8.6	169915	2	AC068290	AC068290 Homo sapi

ALIGNMENTS

RESULT 1

AX207624 AX207624 2930 bp DNA linear PAT 31-AUG-2001

LOCUS Sequence 1 from Patent WO0156598.

DEFINITION AX207624

ACCESSION AX207624

VERSION AX207624.1 GI:15422329

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2930)

AUTHORS Hellergvist,C.G.

TITLE Methods for preventing or attenuating pathoangitogenic conditions by using the gbs-toxin (cm101) receptor as a vaccine

JOURNAL Patent: WO 0156598-A 1 09-AUG-2001;

FEATURES VANDERBILT UNIVERSITY (US)

Location/Qualifiers

Source 1..2930

/organism="Homo sapiens"

/db_xref="taxon:9606"

263..1873

/note="unnamed protein product"

```
/codon_start=1
/protein_id="CAC60358.1"
/db_xref="GI:15422330"
/translation="MAAGAMTPRPVQPARPGFGLSGRRSLLCOVASTPAHVGVNRS
PVDLARNDSNTLEDRTPPLPGAPRAEAPVCCSARYNLAIILAFEGFFIVIALRVNLS
VALDMVDNTELEDRTPPLPGAPRAEAPVCCSARYNLAIILAFEGFFIVIALRVNLS
ITIQIPGVASKIGKMLGILGTAIVLFTPIAADLGVGLIVLRALGELGV
TFPAMHAWSSNAPPLERSKLLISVAGAOGLTVISLPLSGIICVYMNWTVYVFPFGT
IGLFWLLMIWLVSDPPOKHKRISHYEXEYILSSLRNOLSSOKSPVWPVILASLPLWA
IYVAHESYNWTFYLLTLLPTYMKEILRNVOENGFLSLPYLGLSWLMLSGQAADN
LRKAMNCTVRRIFSLGIMGPVAFVLAAGFICDYSIAVAFITISTLLGGFCSSG
FSINHLIDFAPSAGVILLGNTFATIPGMVGPVIAKSLTPDNTVGEWQTVFYIAAAN
VFAGIIFTLFAKEGVONWALNDHGHHRH"
BASE COUNT      729 a   698 c   681 g   822 t
ORIGIN
```

Query Match 100.0%; Score 2930; DB 6; Length 2930;

Best local similarity 100.0%; Pred. No. 0;

Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 qtcggtcgaagccctcccttaattatgtgaattcaagtcaccactgcccgccgcaa 60
Db 1 GTTCGGTGAAGCCCTCCCTTAATATGTGCAATTCAGTCCCCACGTGCCCGCCGCAA 60
Qy 61 gcccactcatctcgtcgtcggcaggggtggccctgcactttacaagggggtgcagga 120
Db 61 GCCCCCACTCATCTCGTCTGCGGCGAGGTGCGCCCTGCACATTACAGGGGTGCAGGA 120
Qy 121 gcgggagacggtcgtccgaacacggctcccgccgcatgqtagaccggggcgagcgg 180
Db 121 GCGGGAGACGCTGTCGGAACACGGCTCCCGGCATGCGTAGACCGCGCGGCGAGCGG 180
Qy 181 gctcaattgcacacacacacacacacacacacacacacacacacacacacacac 240
Db 181 GCTCACTTTGCGGCAATCCTACGAGAACCTCCAGAACCTCGCGTTCCTAGTCCAAACCCAA 240
Qy 241 gccagagtgccacacacacacacacacacacacacacacacacacacacacac 300
Db 241 GCCAGAGTTGCCACACCTTAAGATGCGCGGCGGCGGCGATGACACCGCCCGCGTCCA 300
Qy 301 gccagctcggccgggggttcgggttcggggccggcgcctcctcctcctcctcctc 360
Db 301 GCCAGCTCGGCCCGGGGCTTCGGGCTGTCGGGCGCGCGCTCCTCTCTGCGAGGTGGC 360
Qy 361 gactacacgtcagctagcgtcagctcagctcagctcagctcagctcagctcagctc 420
Db 361 GACTACACCTGCTCAGTAGCGCTCATGAGGTCTCCGGTTCGAGACCTGGCCCGGAACGA 420
Qy 421 tggcaggagacacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 480
Db 421 TGGCGAGGAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 480
Qy 481 tcagtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 540
Db 481 TCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 tgtgtatgcatcagctgcatcagctgcatcagctgcatcagctgcatcagctgcatc 600
Db 541 TGTGTATGCAATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 601 aactttagaagataatagaactccaagcgctgtccagagcattctgctcccaataaag 660
Db 601 AACTTTAGAAGATAATAGAACCTTCCAAGCGCTGTCCAGAGCATTCCTCCCATATAAAGT 660
Qy 661 tcatcataataacacgggtaagaagtaccatgggagtcagaaactcaaggatcgattct 720
Db 661 TCATCATATAATCAACCGGTAAGAAGTACCAATGGGATGCAGAAACTCAAGGATGGATTCT 720
Qy 721 cgggtcctttttttagctacatcatcacagattcctcgtgaggtatgtgcccagcaa 780
Db 721 CGGTTCTCTTTTATGGCTACATCATACACAGATTCCTGGAGGATATGTGGCCAGCAA 780
```

```
Qy 781 aataggggggaaaaactgctgtaggatttgggacaccttggcactgctgctcaccctgtt 840
Db 781 AATAGGGGGGAAAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGTCTCACCCTGTT 840
Qy 841 cactccactgctgcagatttaggagttggaacctcatttactcagagcactagaagg 900
Db 841 CACTCCCAATGCTGCAGATTTAGGAGTTGGACCTCATTTGTAAGTACTCAGAGCCTAGAAG 900
Qy 901 actagagaggggtttacatttccacgaatgacatgcatgctggtctcttctggctcccc 960
Db 901 ACTAGGAGAGGGGTGTACATTTCCAGCCATGATGCATGTGGTCTCTCTTGGGCTCCCC 960
Qy 961 tcttgaaagaagcaaaactctttagcatttcgtatgcagagacacagctgggacagtaat 1020
Db 961 TCTTGAAGAAGCAAACTCTTAGCATTTCTGATGATGATGATGATGATGATGATGATGAT 1020
Qy 1021 tctctctctctctctctctctctctctctctctctctctctctctctctctctctct 1080
Db 1021 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Qy 1081 ttttgcactatttgaatttttgggttcttcttggatcctggttagttagtgaacac 1140
Db 1081 TTTTGTACTATTGGAATATTTTGGTTCTTTTGTGGATCTGGTTAGTGTAGTGAACACC 1140
Qy 1141 acaaaacacagagaaatttccattatgaaaaggaaatcacattcttctcatcattaagaaa 1200
Db 1141 ACAAAAACACAGAGAAATTTCCCATTTATGAAAAGGAATACATTTCTTTCATCATTAAGAAA 1200
Qy 1201 tcagcttcttccacagaagtcagtcggtgggtaccctattttaaataccctgcacattt 1260
Db 1201 TCAGCTTTCTTTCACAGAAAGTCAGTCGCGTGGGTACCATTTTAAAAATCGCTGCACATTG 1260
Qy 1261 gactatcgtagtgcacactttcttcaactggacacttttatactttatactttatactatt 1320
Db 1261 GGTATCGTAGTGTGCACACTTTCTTACAACTGGACATTTTATATACATTTATGACATATT 1320
Qy 1321 gctacttatatgaaggagatcctaaggttcaatgtcaagagaatgggttttttctctc 1380
Db 1321 GCCTACTTATATGAAGGAGATCCTAAGGTTCAATGTTCAAGAGAATGGGTTTTTATCTTC 1380
Qy 1381 attgcctatttaggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
Db 1381 ATTGCCTTATTTAGGCTCTTGGTTATGTATGATCTCTGTCTGTCAGAGCTGCTGACAAATTT 1440
Qy 1441 aaggggcaaatgaaatttcaactttatgttgcagaaatttttagccttatagaagaat 1500
Db 1441 AAGGGCAAAATGGAATTTTCACTTTATGTGTGCCAGAAATTTTAGGCTTTATAGGAAT 1500
Qy 1501 gattggacctgcagtatctcctggtgagctgctggttcattgctgtgattattcttggc 1560
Db 1501 GATTGGACCTGCAGTATTCTGCTAGTCTGCTGGCTTCATTGGCTGTGATTTATTTCTTGGC 1560
Qy 1561 cgttcttctcctaatatatacaaacactggaggcttttctctctctctctctctctctct 1620
Db 1561 CGTGTCTTTCTTAACATATATCAACAACTGGGAGGCTTTTGTCTCTCTGGAATTTAGCAT 1620
Qy 1621 caaccaatcggtattgctcctcctgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
Db 1621 CAACCAATCGGTATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1680
Qy 1681 tggcactattccaggaatggttggcccgctcattgtctaaagtctgacccctgataaac 1740
Db 1681 TGGCCTATTTCAGGAATGTTGGGCCGCTCATTTGTAAGGCTGACCCCTGATAAACAT 1740
Qy 1741 tgttgagaaatggcaaacccgtgttctataattgctgctgtattattgttttgggtgc 1800
Db 1741 TGTTGAGAAATGGCAACCGCTGTTTATATTGCTGCTGTATTAATGTTTGGTGGCCAT 1800
Qy 1801 tttcttatacattatcgccaaagggtgaagtacaaaactgggctctcaatgatcaccatgg 1860
Db 1801 TTTCTTACACTATTCGCCAAAGGTGAAGTACAAAACCTGGGCTCTCAATGATCACCATGG 1860
Qy 1861 acacagacactgaaggaaacaaataaataactgctctatttaattattttatttattc 1920
```

```
Db 1861 ACACAGACACTGAAGAACCAATAAATAATCTCCCTCTATTATATGATATTTATTATATC 1920
QY 1921 atgaacctcaagtcctctgtattgtgtaagcattctatgtctcttttttaattgtac 1980
Db 1921 ATGTAACTCAAGTCCCTCTGTATGTGTAGCAATCTATGTCTCTTTTAAATGTATC 1980
QY 1981 ttgtattagatttttaagcctataatcatgaataatcatcactagttgccagaataataaaa 2040
Db 1981 TTGTATTAGATTTTAAAGCCCTAATCATGAATAATCACTACTTCCAGAAATAATAAAA 2040
QY 2041 tgaactgtgttaattataataatgttaagctaggaactctcacttttaggttcaacatc 2100
Db 2041 TGAACCTGTCTTAAATATGAATGAATGAAGCTAGGACTTCTACTTTAGGTTTCACATAC 2100
QY 2101 ctgctctgactgcggcaacatgaatagacagttctctgtatttttttagggccatact 2160
Db 2101 CTGCCCTGTAGTCGGCAACATGAATAGGACAGTCTGTGTATTTTAAAGGCCATATC 2160
QY 2161 aaaggaatgagctgaacagacctctgatacctttgtcttaattaaacttagatgataat 2220
Db 2161 AAAGGAATGAGCTGAACAGACCTCTGTATACCTTTGCTTAATTAACCTAGATGATAAT 2220
QY 2221 tctcaggtactgataaacacctgtgtgttctcactttctctcctcctcagctctc 2280
Db 2221 TCTCAGGTACTGATAAACACCTGTGTGTCTACTTTCTCTCATATAAAATTTGTGAGCTCTC 2280
QY 2281 tctgacacttagacctcaactttagcatctctgtgagctgacctccactgtataattt 2340
Db 2281 TCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGAGCTGCCATCCACTGTATTAATTT 2340
QY 2341 cgctggcaactggactgaggaggagtggtgccaggagctgccaagcactccctccctgg 2400
Db 2341 CGCCTGGCAACTGGACTGAGGGGAGTGTGCCAGGAGCTGCCAAGCACTTCCCTCCCTGG 2400
QY 2401 ctccaggttcagagtgccagcgtttatcagagcagcagcattccaaagccagagccagtgc 2460
Db 2401 CTTTCAGGGTCAGAGTCCCGACGCTTTATCAGAGGCAAGCATCCAAAGCCGAGCACTGTCTC 2460
QY 2461 gactctcggtctggtccttctctgagggggtatcaatgtgtagataaagccctgagt 2520
Db 2461 GACTCTTCGGCTGGTGCCTTCTCTGAGGGGTATCAATGTGTAGATAAAGCCCTGAGT 2520
QY 2521 aggaagagcagtgatgatacactgctatggtcttgatatacctcctcactccctccc 2580
Db 2521 AGGCAAGAGCAGTGAGATCCACTGCTATGCTCTTGATACATCTCCTCAAACTTCTCCCTCCC 2580
QY 2581 agcaagaggaatattggtggaatgcaacctgcacaaagaaataatgcgaagcgccgggc 2640
Db 2581 AGCACAGAGGAATATTGGTGGCATCAACCTGCAAAAGAAATAATCGAAGCGCGCGGC 2640
QY 2641 acggtgctcatgctgtaatccagcactttggggggtgaggtggcggaatcatgaga 2700
Db 2641 ACGGTGCTCATGCTCTGTAAATCCACGACTTTGGGGGGCTGAGTGGCGGAATCATGAGA 2700
QY 2701 tcaggagttcgagaccagcgtggccagcagtggtgaaaccccatctctactataaaatacaa 2760
Db 2701 TCAGGAGTTTCGAGACAGACCGCTGGCCAGCATGTTGAAACCCCATCTCTACTATAAAATACAA 2760
QY 2761 aaaaatagctggcgtggtgacgggcgcctgtaattcccagatactcaggaggtgaggtga 2820
Db 2761 AAAAATTAGCTGGCGGTGGTGACGGGGCCCTGTAAATCCCAGATACTCAGGAGGTGAGGTA 2820
QY 2821 ggagaatcactgaaacctggagggtggaagtgcagtgaaaccaagatcacgcgcaactgcac 2880
Db 2821 GGAGATCATCTGAACCTGGGAGGTGGAAGTGTGCACTGAACCAAGATCATCGCCACTGCAC 2880
QY 2881 tccagcctgggagatggagcagactcccaactcaaaaaaataaaaaa 2930
Db 2881 TCCAGCCTGGCGCATGGAGCGAGACTCCAACCTCAAAAAAATAAAAAA 2930
```

RESULT 2

```
AF244577
LOCUS Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
DEFINITION Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
ACCESSION AF244577
VERSION AF244577.1 GI:9719373
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2930)
AUTHORS Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P.,
Carter,C.E., Shi,E., Venkov,C., Yakes,M.F., Page,D.L. and
Hellerqvist,C.G.
TITLE Identification of a novel membrane protein from mammalian cells
that interacts with the anti-pathoangiogenic compound CM101
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2930)
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School
of Medicine, 23rd Pierce, Nashville, TN 37232-0146, USA
FEATURES
Source
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2930
/gene="HP59"
CDS
263..1873
/gene="HP59"
/note="similar to Na+ inorganic phosphate transporter"
/codon_start=1
/product="membrane glycoprotein HP59"
/protein_id="AAF97769.1"
/db_xref="GI:9719374"
/translation="MAGAMTPRPVQPARPGFGLSGRRSLLCQVASTPAHVGMVRS
PVRLARNDGEESTDRTPLLPGRAEAPVCCSARYNLAILAFGFFIVIALRVNLS
VALVDMVDSNTTLEDNRTSKACPSHSAPIKVHNHTQTKYQWMDAQEQWILGSLFYGY
ITQTIPGGVASKIGKMLLGFILGTAVLTFTIAADLGVGLPIVLRALGLGEGV
TPPAMHMSWAPLERSKLLISVAGAQLGTVISLPLSGIICYVMWTVYVFPGT
EIFWFLWILWLVSDTPQKHRIKSHVEKEYILSLRNOLSSOKSPVWPIKLSPLWA
IVAFNSYNWFYTLTLTLPYMKELIRENQENGFLSLPYLGSWLCMILSQAADN
LRAKNFSLCVRRIIFSLIGMIGPAVLVAGFICGDSLAVAFITSTITLGGFCSSG
FSINHLDIAPSYAGILLTGTTFATPGMGPVIAKSLTPDNTVGEMQTVFYIAAII
VFGATFFTLFAKGEVQNNALNDHGHRRH"
BASE COUNT 729 a 698 c 681 g 822 t
ORIGIN
```

```
Query Match 100.0%; Score 2930; DB 9; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgtaagacctcccttaattatgtgcaattcaagtcacctgcccacccgcaaa 60
|||||
Db 1 GTTCGTGCGAAGCCCTCCCTTTAATTATGTGCAATTCAGTCCCTCCCTCCCGCAA 60
|||||

QY 61 gcccccactcatctcgctgcgggcaggtgcccctgcactttacaaggggtgcagga 120
|||||
Db 61 GCCCCCACTCATCTCGCTGCGGCGAGGTTGCCCTTGACACTTTACAAGGGGTGCAGGA 120
|||||

QY 121 gggggagacggtgctccgaacacccggtcccccggcatgcgtagaccggtgcggcgagcg 180
|||||
Db 121 GCGGGAGACGGTCTGCCGAACACCGGCTCCCGGCATGCGTAGACCGCGGGCGGAGCGG 180
|||||

QY 181 gctcaatttcgccaatcctacagaaactcccaagaactccggttccctagtcacacccaa 240
|||||
Db 181 GCTCACTTTGCGCAATCTCTACGAGAACTCCAGAACTCCGCTTCCTAGTCAACCCAA 240
|||||

QY 241 gccagagttgccacacacctaaagatggcgggggtgcgatgacaccccccgcggtcca 300
|||||
Db 241 GCCAGAGTTGCCCAACACTTAAGATGGCGGGGGGGCGATGACACCCCGCCCGCGGTCCA 300
|||||
```


Db 2461 GACTCTTCGGCTGGTGCCTTCTCTAGAGGGGCTAATGTTAGATAGAAAGCCCTGAGT 2520
QY 2521 agcacaagacagtgagatccactgctatgcttgatgatacatctcaacttccctccc 2580
Db 2521 AGGCAGAGACAGTGAATCCACTGCTATGCTTGTGATACATCTCAAACTTCCCTTCCC 2580
QY 2581 agcacagagaatattggtgcatgcaacacctgcacaaagaaatgcgaagcgccgggc 2640
Db 2581 AGCACAGAGAAATATTGGCTGGCATCAACCTGCACAAAGAAATGCGAAGCGCGCGGC 2640
QY 2641 acggtgctcatgctgtaatccacacactttgggggctgaggtgagcgaatcatgaga 2700
Db 2641 ACGTGGCTCATGCTGTATATCCAGCACTTTGGGGGGCTGAGGTGGCGAATCATGAGA 2700
QY 2701 tcaggagttcgagacagcctggcagcagctgtgaaaccccatcttactataaaatacaa 2760
Db 2701 TCAGGAGTTCGAGACAGCCTGGCAGCATGTTGAAACCCCATCTCTACTAATAATACAA 2760
QY 2761 aaaaattagctggcgtggtgacggggcgctgtaatccccagatactcaggaggtcaggta 2820
Db 2761 AAAAAATTAGCTGGCGTGTGTCAGCGGCGCTGTAATCCAGATACTCAGGAGGCTGAGGTA 2820
QY 2821 gagaatacactgaacctggaggtggaagtgcagtggaacaaagatacgcgaactgac 2880
Db 2821 GGAGAACTACTTGAACCTGGGAGGTGGAAGTTGCACTGAACCAAGATCACGCCACTGCAC 2880
QY 2881 tcagcctggcgatggagcagactcccaactcaaaaaaaataaaaaa 2930
Db 2881 TCCAGCCTGGCGATGGAGCAGACTCCAACCTCAAAAAAATAAAAAA 2930

RESULT 3
LOCUS BC020961 3329 bp mRNA linear PRI 22-JAN-2002
DEFINITION Homo sapiens, solute carrier family 17 (anion/sugar transporter), member 5, clone MGC:8885 IMAGE:3847279, mRNA, complete cds.
ACCESSION BC020961
VERSION BC020961.1 GI:18088702
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3329)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdépxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAC Plate: 20 Row: 1 Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6912665.
Location/Qualifiers
1..3329
/organism="Homo sapiens"
/db_xref="LocusID:26503"
/db_xref="taxon:9606"

/clone="MGC:8885 IMAGE:3847279"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: PCMV-SPT6"
125..1612
CDS
/codon_start=1
/product="solute carrier family 17 (anion/sugar transporter), member 5"
/protein_id="AAH20961.1"
/db_xref="GI:18088703"
/translation="MRSPVRLARNDEESTDRTPLLPGAPRAEAPVCCSARYNLAI
LAFGFFIYALRVNLVSALVMDVNTLEDNRTSKACPEHSAPIKVHHNOTKKYQ
WDAETQWILGSEFFGYITIPGQYVASKIGKMLGFLGTAVLFTPTIADLG
VGPLIVLRALGLEGGVTFPAMHAMWSWAPPLERSKLLSISYAGOLGTVISLP
LSC IICVYMNWTVYVFFGTIGIFWLLIWLVDTPQKHKRI SHYEKEYILSLRNOLSS
QKSPVWPILAKSLPLWAIWVAHFSYNTFTYLLLLPYMKELIFNVQENGFLSSLP
YLGWMLGSAADNARWNSFCRRIFSLIGMIGPAVLEAFVAGGFTIGCDLSLA
VAFITISTLGGFCSSGFSINHLDIAPSYAGILGITTTFATIPGMGVPVIAKSLTPD
NTVGEWQTVFYIAAANVGAIFFTFLFAKGEVQWALNDHGHHRH"
BASE COUNT 863 a 718 c 727 g 1021 t
ORIGIN
Query Match 89.4%; Score 2619.2; DB 9; Length 3329;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2627; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 290 cgcgcgggtccagccagctcgcccgccgggttcgggctcggcgccgctccctctc 349
Db 29 CCCACGCGTCCGGCGCGCTCGGGCGGCTGTCGGGCGCGCGCTTCCCTTCTC 88
QY 350 tgccaggtggcagtcacactgctcacgttagcgtcagtcaggtctccggttcgagacctg 409
Db 89 TGCCAGGTGGCCAGTACACTGCTCACGTAGCGGTGATGAGTCTCCGGTTCGAGACCTG 148
QY 410 gcccggaacgctggcgagagagacggacggacgcacgcctctctaccggcgcccaagg 469
Db 149 GCCCGGACGATGGCGAGGAGACGCGACGCGCTCTTCTACCGGGCGCGCCACGG 208
QY 470 gccgaagccgctccagtgctgctcgtctcgtctcaacttagcaatttggcctttt 529
Db 209 GCCGAAGCCGCTCCAGTGTGCTGCTGCTGTACAACTTAGCAATTTGGCCCTTTT 268
QY 530 ggtttcttcattgctgattacgttgatcgtggaatcgtgagtggttgctgagtgatggtta 589
Db 269 GGTTCCTCTATTGTGTATGATACGTTACGTTGATCTGAGTGTGCGTGTAGTATGTTA 328
QY 590 gattcaaatcaacttttagaagataataagaacttccaaagcgctgtccagagcattctgct 649
Db 329 GATTCAAAATACAACTTTAGAAGATAATAGAACTTCCAAAGCATGTCAGAGCATTTGCT 388
QY 650 ccataaaagtcatcataataataacacgggtaagaagtaaccatggatggatgagaactcaa 709
Db 389 CCATATAAAAGTTTCATCATATCAACCGGGTAAGAAAGTACCAATGGGATGCAGAAATCAA 448
QY 710 ggatgattctcggttccttttttatgctacatcatcacagattctctgagagatat 769
Db 449 GGATGGATTCTCGGTTCCTTTTATGGCTACATCATCACAGATTCCTGGAGGATAT 508
QY 770 gttgcagcaaaaataggggggaaaaatgctgtaggatttgggattcttggcactgctgctc 829
Db 509 GTTGCCAGCAAAATAGGGGGGAAATGCTGCTAGGATTTGGGATCTCTGGCAGTGTGTC 568
QY 830 ctcaacctgttcaactccattgctgcagatttagaggttggaccactaatgttactcaga 889
Db 569 CTCACCTCTTCACTCCCATTTGCTGCAGATTTAGGAGTTGGACCACTCATTTCTACTCAG 628
QY 890 gcactagaaggactaggagaggtgttacatttcagccatgcatgcatggtctctct 949
Db 629 GCATAGAAAGGACTAGGAGAGGGGTGTACATTTCCAGCCATGCATGCCATGTGGTCTCT 688
QY 950 tggggtccccctctctgaaagaagcaaaccttcttagcatttcgtatgcaggagcacagctt 1009

QY 2101 ctgcctctagtcggtcggaacatgaagtagcacagttctctgttgatttttagggccatact 2160
|||||
Db 1986 CTGCTCTCTAGTCGGCAACATGAGTAGCACAGTCTCTGTGATTTTATAGGCCATACT 2045
|||||
QY 2161 aaaggggaatgagctgaacagacctcctgatacctttgcttaataaactagatgataat 2220
|||||
Db 2046 AAAGGGAATGAGCTGAACACAGCTCTGATACCTTTGCTTAATAAATAGATGATAAT 2105
|||||
QY 2221 tctcaggtactgataaacacctgttgttcttaactttccataaaaaattgtcagctctc 2280
|||||
Db 2106 TCTCAGGTACTGATAAACACCTGTTGTTGTTCACTTTCCTCATAAAAATGTTCAGCTCTC 2165
|||||
QY 2281 tctgacacttagacctcaaaactttagcatctctgtgagctgcctccactcgtataattt 2340
|||||
Db 2166 TCTGACACTTAGACCTCAAACTTTAGCACTCTCTGTGAGCTGCCATCCACTGTAATAAT 2225
|||||
QY 2341 cgctggcaactggactggaggggagtggtgccaggcagctgccaagcaactccctccctgg 2400
|||||
Db 2226 CGCTTGCAACTGGACTGAGGGAGTGTCGCCAGGCAGCTGCCAAGCACTCCCTCCCTGG 2285
|||||
QY 2401 ctccagggctcagatgccagctttatcagagcagcatccaagccagagccagctgtc 2460
|||||
Db 2286 CTTCAGGGTCAGAGTGCCCAAGCGTTTATCAGAGGCAGCATCCAAGGCCAGCCAGTGTC 2345
|||||
QY 2461 gactcttcggctggtgctcttcctctgaggggctatcaatgtgtagataaagccctgagt 2520
|||||
Db 2346 GACTCTTCGGCTGTCCTTCCTCTGAGGGCTATCAATGTGTAGATAAAGCCCTGAGT 2405
|||||
QY 2521 agcagagcagtgagatccactgctatggtcttgatatacctcaaaatttccctccccc 2580
|||||
Db 2406 AGCCAGAGCAGTGAGATCCACTGCTATGCTCTTGATACATCTCAAACTTTCCCTTCC 2465
|||||
QY 2581 agcacagagaaatctggctggcatcaacctgcacaaagaaataatgc 2627
|||||
Db 2466 AGCACAGAGGAATATGGCTGGCATGCACCTGCAAAAGAAAATGC 2512
|||||

RESULT 5
HSA387747
LOCUS
DEFINITION Homo sapiens mRNA for sialin.
ACCESSION AJ387747
VERSION AJ387747.1 GI:6562532
KEYWORDS sialin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2512)
Verheijen,F.W., Verbeek,E., Aula,N., Beerens,C.E., Havelaar,A.C.,
Joosse,M., Peltonen,L., Aula,P., Galjaard,H., van der Spek,P.J. and
Mancini,G.M.
A new gene, encoding an anion transporter, is mutated in sialic
acid storage diseases
Nat. Genet. 23 (4), 462-465 (1999)
20047778
2 (bases 1 to 2512)
Verheijen,F.W.
Direct Submission
Submitted (03-JUN-1999) F.W. Verheijen, Erasmus University, Dept.
Clinical Genetics, P.O. Box 1738, 3000 DR Rotterdam, NETHERLANDS
Location/Qualifiers
FEATURES
1..2512
/organism="Homo sapiens"
/db_xref="taxon:9606"
271..1758
/codon_start=1
/product="sialin"
/protein_id="CAB62540.1"
/db_xref="GI:6562532"
/translation="MRSPPVRLARNDGESTDRTPLLPGAPRAEAPVCCSARNLAI
LAFFGFFIVIALRVNLVALVDMVDSNTTLEDNRTSKACPEHSAPIKVHHNOTGKKYQ

WDAETQGWILGSFFYYIITQIPGGYVASKIGGKMLGILGTAVLFLFTPIADLG
VGLTVLRALGIGBVTFFPAMHMSWAPPLERSKLLSISYGAOLGTWISLPISG
IICYYMNTYVYFEGTIGIFWFLIWLIVSDTPQKHRI SHYEXEYTLISLRNOLSS
OKSPVWPILASLPLALVVAHFSYNTFFYLLPILPYMKELLPFNQENGFLSSLP
YLGWLMILSGQAADNRKWNFTSLCVRRIFFSLIGMIGFAVLVAAGFLGCDISLA
VAFITISTLLGFCSSFSINHLDIAPSYAGILLGTTNTFATIPGMGPVIAKSLTPD
NTVGEWQTVFYIAAAINVFAGIFFTLFAKEGVQWALNDHHRH"
BASE COUNT 612 a 585 c 566 g 749 t
ORIGIN
Query Match 80.5%; Score 2359.4; DB 9; Length 2512;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2437; Conservative 0; Mismatches 6; Indels 64; Gaps 1;
QY 185 actttgcgcaactccttaagacactcccaagacactccgctccctcctagtcacaaacccagca 244
|||||
Db 6 ACTTTGCGCAATCTTACGAGAACTCCCAAGAACTCCGCTTCCTCTAGTCCAAACCAAGCA 65
|||||
QY 245 gqgttgccacacacctaagatgcgcgggg-----cgatgacacccgccccccggtcca 275
|||||
Db 66 GAGTTGCCACACCTAAGATGCGGGGGGGGGGGGCGCGCGCGCGCCCTCTGGGGCG 125
|||||
QY 276 -----cgatgacacccgccccccggtcca 300
|||||
Db 126 GACCGCGGGACTAGACGTGGCGGGGGCGGTGTCTATCGCCCCCGCCCGCGGTCCA 185
|||||
QY 301 gccagctcgcccgggggttcgggctgctgggcggcgggcctcctctctcgcaggtggc 360
|||||
Db 186 GCCAGCTCGCCCGGGGGCTTCGGGCTGTCGGGGCGGGCGCTCCCTTCTTGCAGGTGGC 245
|||||
QY 361 gqgtacacctgctcagtagcgctcaggtctcgggttcgagacctggcgccgaaagca 420
|||||
Db 246 GAGTACACCTGTCTACGTAGGCGTCTAGAGTCTCGGTTTCGAGACCTGGCCCGGAACA 305
|||||
QY 421 tggcgagagagcagcggaccgcctctctacccggcgccccacagggcgccgaagccgc 480
|||||
Db 306 TGGCGAGGAGAGCAGCGGACCGCACGCTCTTCTACGGGGCGCGCCACGGCGGAAGCGCG 365
|||||
QY 481 tccagtgctgctctgctcgttacaacttagcaatttggccttttttggtttcttcac 540
|||||
Db 366 TCCAGTGTGCTGCTCTGCTGCTGTACAACTTAGCAATTTTGGGCTTTTGGTGTCTTCAT 425
|||||
QY 541 tgttatgcattacgtgtgaatctgagttgctgttagtggtatggttagattcaaatac 600
|||||
Db 426 TGTGTATGCAATTACGTGTGAATCTGAGTGTTCGGTTAGTGATGTTAGATTCAATAC 485
|||||
QY 601 aactttagaagataatagaacttccaaaggcgtgtccagagcattctgtccataaaagt 660
|||||
Db 486 AACTTTAGAAGATAATAGAACTTCCAAGGCGTGTCCAGAGCATTTCTGCTCCCATAAAAGT 545
|||||
QY 661 tcatcataatcaaacgggtaagaactaccaatggagatgcagaacactcaagatgatctc 720
|||||
Db 546 TCATCATATCAACAGGGGTAAAGTACCAATGGGATGCAGAACTCAAGGATGGATTCT 605
|||||
QY 721 cggttccttttttatggctacatcatcacacagattcctgagagatatgttgccagcaa 780
|||||
Db 606 CGGTTCTTTTATGGCTACATCATCACAGATTCCTGGAGGATATGTTGCCAGCAA 665
|||||
QY 781 aaatggggggaaaaatcgtgctaggatttgggatcccttggcactgctgctcaccctgtt 840
|||||
Db 666 AATAGGGGGGAAAAATGCTGTAGGATTTGGGATCCTTGGCACTGCTGCTCACCTCTGT 725
|||||
QY 841 cactcccatgctcagatttaggattggaccactcattgtactcagacactagaagg 900
|||||
Db 726 CACTCCCATGTCTGAGATTTAGGAGTTGGACCACTCATTTGACTCAGAGCACTAGAAG 785
|||||
QY 901 actaggagagggtgttacatttccagccatgcattccattggtctcttgggtccccc 960
|||||
Db 786 ACTAGGAGAGGGTGTGTACATTTCCAGCATGCATGCCATGTGCTTCTTCTGGGCTCCCC 845
|||||
QY 961 tcttgaagaagaacaaactctttagcatttctgatatcagagagcacagcttggagacagtaat 1020
|||||

/product="membrane glycoprotein SP55"

/protein_id="AAF97770.1"

/db_xref="GI:9719376"

/translation="MKSPVSDIAPSDGEGSDRTPLLQRPRAEPAPVCCSARYNLAF
LSFFGFVLYSRVNLVALVDMSNTAKDNRTSYCAEHSAPFKVLHNOGTGKKYR
WDAEQFTGLSGFFGYITITPGYVASRSGKLLILGFGIFATIFLFTPLAADFG
VGLVALRLEGIGEVTPAMHAMWSWAPLERSKLISISYAGAQLGTVVSLPLSG
VICYYMNWTVYFVFGIGVIWFIWILWICLVSDTPETHKITITPYEKYILSLKLNOLSS
QKSPWIPMLKSLPLWALVAVHFSYWFYTLTLPLPYMKLEVLRFNTOENGFLSAPV
YIGWLCHLSQADNLRLRNWFSFLWVRVFSLIGMIGFPAIFLVMAGFICGDDYSLA
VAFLLTISLGLSCSSGFSINHLDIAPSAGILGLITNTFATIPFMIGIPIIARSLTPE
NPIGEMQVFCIAAANVFGAIFLFLFARKEQVNAISDHQHRN"

BASE COUNT 745 a 623 c 594 g 882 t

ORIGIN

Query Match		53.6%;	Score 1570.8;	DB 4;	Length 2844;
Best Local Similarity		82.0%;	Pred. No. 0;		
Matches 1916;		Conservative 0;	Mismatches 387;	Indels 33;	Gaps 8;
QY	309	gccccggggcgtcgagctgcggccgcgcgtccctctctgccaggtggcgagtcacac	368		
Db	5	GGGGGGGGGGCTCGCGGCTCCCGTGGAGCTCTCTTTCCCGGAGAGAGTTTCGGCC	64		
QY	369	ct--getcacgtgagcgtcatgaggtctccggttcgagacctggcccgcgaacgatgcga	426		
Db	65	GTAGCTCCCTGAAGGCATCATGAAGTCCCGGTTTCGGAGTTAGCCCGGAGCGACGCGA	124		
QY	427	ggagagcaggaacgcacgcctctcttacggggcgcccccaacggggcggaagcgccacgt	486		
Db	125	GGAGGGCTCGGACCGCACACCGCTCTGTGAGCGCGCCCGCGGGGAAACCCGCTCCAGT	184		
QY	487	gtcgtcctcgtcgttacaaacttagcaatttggcctttttggtttctctcattgtga	546		
Db	185	ATCCTGCTGTGCTCGTTACNACCTACGATTTTGTCTTTTGGTTTCTTCTGTTCTCTA	244		
QY	547	tgcattacgtgtgaatctcgtggtgttcgttagtggataggttagattcacaatacaacttt	606		
Db	245	TTCAATACGGGTGAATCTGAGCGTTGCACCTAGTGGACATGGTGGATTCAAAACAACATGC	304		
QY	607	agaagataagaaactccaagcggtgtccagagcattctgtcccaataaaagattcatca	666		
Db	305	CAAAGATAATAGAACGTCTTACGAGTGTGCAGAGCAATCTGCTCCCATAAAGTCTTCA	364		
QY	667	taataaacgggttaagaagtaccaaatgggatgcagaaactcaaggatggattctcggttc	726		
Db	365	CAACCAACGGGTAAAGATACCGGTGGGATCGAGAACTCAAGATGATTCGCGATC	424		
QY	727	ctttttttatggtacatcatcacagatctcctggaggatatggtggcagcaaaatagg	786		
Db	425	TTTTTTCTATGGCTACATCATCACAAAATTCCTGGAGGATATGTTGCCAGCAGAAGTGG	484		
QY	787	ggggaaaatgctcgtaggatttggatctcttggcactgctgtcctcaacctgttcaactcc	846		
Db	485	GGGGAAGCTGTGCTAGGATTCGGGATCTTTGTGTCAGAGTATCTTCACCCCTGTACTCC	544		
QY	847	cattgctgcagattttaggatttgaccactcattgtactcagagcactagaagactagg	906		
Db	545	CCTCGCTGCAGATTTTCGGAGTCGGAGCCCTTGTTCAGCTCAGGGCACTAGAAGGGCTAG	604		
QY	907	agaggggtttaacttccagccatgcattgccattggtgtctcttcttggcctcccccttga	966		
Db	605	AGAGGGTGTACATATCCAGCCATGTCATGCCATGTGGTCTTCATGGGCTCCCCCTCTTGA	664		
QY	967	aaqaacaaactctcttagcattctcgtatgcaggagacacagcttggagacagtaattctct	1026		
Db	665	AAGAACAAGCTTCTGAGTATTTTCATATCAGGAGCACAACCTTGGGACAGTAGTTTCTCT	724		
QY	1027	tctcttcttggaataattgctactatatgaattgacctatgcttcttactctttttgg	1086		
Db	725	TCCTCTTTCTGGAGTAATTTGCTACTATATGAATTCGACCTATGCTCTCTATTTCTTGG	784		
QY	1087	tactattggaatatatttggttctcttcttggatctcgtggttagttagtgacacacacaaaa	1146		

Db	785	CATTGTTGGAATCATCTGCTGTTATTTTATGGATCTGCTTAGTTAGTATACACCGAAGAC	844		
QY	1147	acacaagagaattcccatatgaaaaagaaatacatctcttcatcatttaagaataacagct	1206		
Db	845	TCACRAGACAACTCACTCCCTATGAAAGGAGTATATCTTTCATCATTTAAAAAATCAGCT	904		
QY	1207	tcttcacagaagtcagtcgctgggtacccatttttaaatacccgcgcaacttggcgtat	1266		
Db	905	CHCTTCACAGAAGTCAGTCCCGTGGATACCTATGCTGAATACTCACTGCCACTTTGGGCTAT	964		
QY	1267	cgtagttgcacactttcttcaaacctggactttttatacttattgacattatttgcctac	1326		
Db	965	TGTCGTTGCACATTTTCTTACAACCTGGACTTTTATATACTTTGTTGACCTTATTCGCTAC	1024		
QY	1327	ttatatgaaggagatcctcaagtccaatgttcaagaagaatgggttttttctcttcattgc	1386		
Db	1025	TTACATGAAGGAAGTCTCAAGGTTCAATATTCAAGAGAATGGGTTTATCTGCGATGCC	1084		
QY	1387	ttatttagcctcttggttatgatcctgctcaagctgctgcaagctgctgacaatttaagggc	1446		
Db	1085	TTATTTAGGTTGTTGGTTATGATGATCTCTGCGGTCGAAGCTGCTGACAAATTTAAGGGC	1144		
QY	1447	aaaatggaaatttttcaactttatgtgttcgcagaaatttttagccttataggaaatgattgg	1506		
Db	1145	AAGATGGAATTTTCAACTCTCTGGGTTGCAAGAGTTTATAGCTGTGATTTATAGGATGATTGG	1204		
QY	1507	acctgcagttattccgtgtagctgctggtctcattggtctgatttcttcttgcgcttgc	1566		
Db	1205	ACCTGGGATATTCCTGTTTGGCGCAGGATTTATAGGCTGTGATTTATCTCTGGCTGTGC	1264		
QY	1567	tttcttaactataatacaaacactgggaggttttgcctctctggtatttagatcaaca	1626		
Db	1265	ATTCCTAACCATATCAACAACCTCGGAGGCTTTTGCTCTCTGGATTTAGCATCAACCA	1324		
QY	1627	tctggatttgcctctcgtatgctggtatcctcctcctgggcatcaacaataacatttgcac	1686		
Db	1325	TCTGGACATTTGCTCTCGTATGCTGTTATAGGCTGTGATTTATCTCTGGCATACACAAATAC	1384		
QY	1687	tattccagaagaattgggtggcgcgtcatctcaaaagctgacccctgataacacattgg	1746		
Db	1385	TATCTCTGGAATGATTTGGGCCCATCATATGCCAGAAAGCTTACCCCTGAGAACA	1444		
QY	1747	agaatggcaaacgctgttctctatattgctgctgctctatataatgttttgggtgcacatttctt	1806		
Db	1445	AGAATGGCAAACTGTTTCTGTCATCGCTGCTGCTATCAATGATTTTGGTGCCATTTCTT	1504		
QY	1807	tacactattgcacaaaggtgaagtacaaaactgggctctcaatgatcacccatggacacag	1866		
Db	1505	CACACTATTTCGCAAAAGGTGAAGTGCAAAACCTGGGCCATCAGTGCATCACAAGGACACAG	1564		
QY	1867	acactgaaggaaaccaataaataactcctcctcttataatgatttttttatttcatgttaa	1926		
Db	1565	AAACTGAAGGAACCAATAAATAATCTGCTCTATTAATGATATCTTTGTTTATCATGTAA	1624		
QY	1927	ctctaaagtgcctt-----ctgtattgtgaagcatctctatgt-----cttttttt	1972		
Db	1625	CCTAAAAGTGCCTTTTCATATTTTAAATGCTAAGCAATCTATATACAGATAAATAATGTAC	1684		
QY	1973	aattgtacttcttagatttttaaggccctaataatgaataataatgcatttgcagaa	2032		
Db	1685	TAGAAAAATGTGTAGATTTTGAAGGCTTGTAAATCATGAAATGCATAGTAGTGGCATAT	1744		
QY	2033	taataaaatgaactgtgttataattatgaataataatgaactgagacttcttacttaggt	2092		
Db	1745	AAGCAAAATAGCTATTTTAAATATTTATTAACCCCTTTTGCCTGGAACTTACAACTCAGGG	1804		
QY	2093	tcacataccctgctcgtcgtcgggcaacatgaagtaggacagcttctgttatttttagg	2152		
Db	1805	TCACATACTGCTGCAAGTCAGGCAACCCACAATAAGGGAGTCTATTTTA-TTTATAAG	1863		
QY	2153	gccatactaaaggaatgagctgaaacagacctcctgatcaccttctgcttaataaactag	2212		

Db	1864	ACCATACCTAAGAGATGAGCTGAATAGACACCTTCTATACCTTTGCTTAATTAAGGTGG	1923
Qy	2213	atgataattctcaggtactagataaacacacgtgtgtgtgttcaacttctcctaataaaa-----	2267
Db	1924	ATAATAATTCTCAGGTCTTGTAAACATCTGTTTTTTGTACACCTTCTCTCAAAAATTAAT	1983
Qy	2268	-attgtcagctctctctgacacacttagacctcacaacttttagcatctctgtgagctgccat	2326
Db	1984	TGTCATCAGCAATPCCCTGACATGAGTCTCAAACTTTAGCCCTCTCCACGAGCTGGCAG	2043
Qy	2327	ccactgataaattccctggcaactggactgagggagtggtgtccaggcagctgccaag	2386
Db	2044	CCACTGTATCAATTCAGCTGGCAACTTCATGTAGGGAAGCATCCCGAGGAGCTGCCACA	2103
Qy	2387	cactccctccctgtgtccaggttcagagtgccacagcgtttatcatcagaggcagcatccaagc	2446
Db	2104	TGTCCTCTCTGTGCTTCAGGACAGTGCCACCACTT-----AGGCAGCATCCAAGA	2156
Qy	2447	ccaggcaggtgtcgactcttggctgggtggcttctctctctctctctctctctctctctctct	2506
Db	2157	CCAGGCTCAGCGCAAGGCTTTTGACGGTATTCTTCCCTT--GGGGCTGTTAATGTGTGG	2214
Qy	2507	ataaagccctgagtaggaagacagtgagatccactgctctatgtcttgatatacatctca	2566
Db	2215	ATGAAGCCCTGAGCCCAACAGGACAGCGCATCCAGATCATGGTTTCCATGCAACCTC-	2273
Qy	2567	aactttcccttcccagcacagaggaattattggctggcatgcaacctgcaaaagaaa	2622
Db	2274	TCCTTCCCTTCCAGCACACTGAGATATTGCTGCGTATGCTGCAACCTGCAAAAGAAA	2329
RESULT	7		
AX207626			
LOCUS	AX207626	2844 bp	DNA linear PAT 31-AUG-2001
DEFINITION	Sequence 3 from Patent WO0156598.		
ACCESSION	AX207626		
VERSION	AX207626.1	GI:15422331	
KEYWORDS	Ovis sp.		
SOURCE	Ovis sp.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis.		
REFERENCE	1 (bases 1 to 2844)		
AUTHORS	Hellergvist, C.G.		
TITLE	Methods for preventing or attenuating pathoangiogenic conditions by using the gbs-toxin (cm101) receptor as a vaccine		
JOURNAL	Patent: WO 0156598-A 3 09-AUG-2001;		
FEATURES	VANDERBILT UNIVERSITY (US)		
source	Location/Qualifiers		
	1..2844		
	/organism="Ovis sp."		
	/db_xref="taxon:9939"		
CDS	84..1571		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC60359.1"		
	/db_xref="GI:15422332"		
	/translation="MKSPVSDLPSSDGEBSDRTPILORAPRAEPAPVCCSARYNLAF		
	LSFGFFVLYSRVNLVALVDMDSNTAKDNRTSYEAEHSAPIKVLHNGTGYKYR		
	WDAETQWILGFFYGIITQIPGGTVASRSGKLLGLGIFATIFLFTPLADFG		
	VGALVALRLEGEGVTPAMHMSWAPPLERSKLLSIYAGAQLGTVVSLPSLG		
	VICYNNNTVYFFYGIWIWILVICIVSDTPETHKTIPTYREYILSKNLQSS		
	OKSPWIPMLKSLPWLAIWVAHFSYNWTFYLLTLLPTVMKVLRFNIQENGLFSAVP		
	YLGWCLMTLSQADNLRAWNFTLWVRVYFSLIGMTGPAIFLVAAGFICDYSLA		
	VAFUHSITLGGFCSSGSINHLDIAPSYAGILGLITNTFTFPIPMIGFIARSUTPE		
	NTIGBWQVFCIAAINVFGAIFLFLKAGEVQNNALSDHQHRN"		
BASE COUNT	745 a	623 c	594 g 882 t
ORIGIN			

Query Match 53.6%; Score 1570.8; DB 6; Length 2844;
Best Local Similarity 82.0%; Pred. No. 0;

	Matches	1916;	Conservative	0;	Mismatches	387;	Indels	33;	Gaps	8;
Qy	309	ggcccggggcttcgggctcggcgcgcgcctctctccaggtgagagtacac	368							
Db	5	GGGCGGGGGGCTTCGGGGTCCGCTGGAGCTCTCTTTTCCGGGAGCAGGTTTGGGCC	64							
Qy	369	ct--gtcacgtaggcgtcatgatggtctccggttcagacctggcccggaacgatggcga	426							
Db	65	GTAGCTCCCTGAAGGCATCATGAAGTCCCGGTTTCGGACTTAGCCCGGAGCGACGCGA	124							
Qy	427	ggagagcagcagcgcgcgcctctctacggcgccccacagggcggaagccgctccagt	486							
Db	125	GGAGGGCTCGGACCGCACACCCGCTCTGACGCGCCCCCGCGGCGGAACCCCTCCAGT	184							
Qy	487	gtcgtcctcgtcgttacaacttagcaattttggccttttttgggtttctctcattgtga	546							
Db	185	ATGCTGCTGCTCGTGTACAACTAGCATTTTGTCTTTTGGTTTCTTCTGTTCTCTA	244							
Qy	547	tgcattacgtgtgaatctgagtggtgttagtggtatggttagatccaatacaaatctt	606							
Db	245	TTCAATTACGGGTGAATCTGAGCGTTGCACCTAGTGGACATGGTGGATTCAAAACACA	304							
Qy	607	agaagataatagaaactccaagcggtgcagagcattctgtcccaataaaagttcatca	666							
Db	305	CAAAAGATAATAGAACGCTCTACGAGTGTGCAGAGCATTTCTCCCAATAAAAGTCTTCA	364							
Qy	667	taatacaacgggttaagaagtaccaatgggatgcagaaactcaagatggattctcggttc	726							
Db	365	CACCCAAACGGGTAAAAAGTACCGGTGGGATGCAGAACTCAAGGATGGATTCTCGGATC	424							
Qy	727	cttttttttggtcatcatcacagattcctgaggagatctgtgaggatattgttgcagcaaaatagg	786							
Db	425	TTTTTTCTATGCTACATCATCACAAATTCCTGGAGGATATGTTGCCAGCAGAAAGTGG	484							
Qy	787	ggggaaaatgctcaggtattgggataccttggcaactgctgctcaacctgttcaacctc	846							
Db	485	GGGGAAGCTGTGCTAGGATTTCGGGATCTTTGTGACAGCTATCTTCACTGCTGCTACTCC	544							
Qy	847	catgctgcagattaggagttgaccactcattgtactcagagcactagaagactagg	906							
Db	545	CCTCGCTGCAGATTTTCGGAGTCGGAGCCCTTGTTCGACTCAGGGCCTAGAGGGCTAGG	604							
Qy	907	agaggtgtttacatttccagccatgcacgtggtgtgttcttgggtccccctcttga	966							
Db	605	AGAGGGTGTACATATCCAGCCATGCATGCATGTGTGTCTTTCATGGGCTCCCTCTTGA	664							
Qy	967	agaagcaaaactcttagcattcgtatgcagagcacacgttggacagatattctct	1026							
Db	665	AAGAAGCAAGCTTCTGAGTATTTTATATGAGGAGCAGCAACTTGGGACAGTAGTTTCTCT	724							
Qy	1027	tcctcttctgggaataattgctactatatgaattggacttatgctctacttttttg	1086							
Db	725	TCCTCTTCTCGAGTAATTTGCTACTATATGAATTTGACTTATGCTCTTCTATTCTTTGG	784							
Qy	1087	tactatggaataatttgggttcttcttggatcctggttagtagtagtagtagtagtagtag	1146							
Db	785	CATTGTGGAATCATCTGTTTATTTATGATCTGCTTGTAGTGTAGTATACACGAAAC	844							
Qy	1147	acacagagaatttccattatgaaagaatacatcttcttcatatgaagaatacagct	1206							
Db	845	TCACAAGACATCACCTCCCTTATGAAAGGAGTATATCTTTTCATCATTAATAAATCAGCT	904							
Qy	1207	ttctcacagaactgcgctgggtacoccatctttaaataccctgcacatttgggctat	1266							
Db	905	CTCTTCACAGAAGTCAGTCCGTTGGATACCTATGCTGAAATCACTGCCACTTTGGGCTAT	964							
Qy	1267	cgtagttgcacactttcttacaactggactttttatactttattgacattattgactac	1326							
Db	965	TGTCGTGTCACATTTTCTTACAACCTGGACTTTTATACCTTTTGTGACCTTATTGCTAC	1024							
Qy	1327	ttatatgaaggagatcctcaaggttcaatgttcaagagaatgggtttttattcttcattgccc	1386							
Db	1025	TTACATGAAGGAAGTCTTAAGGTTCAATATTCAAGAGAATGGGTTTTTTATCTATCGAGTCCC	1084							

Qy	2447	ccagagccagtgtcgactcttgcgctggctgttcctctcctcaggggctatcaatgtgtag	2506
Db	2157	CCAGGGTCAAGCGCCAAGCTTTGGACGGATATTCCTCCCCT- -GGGCGCTGTAAATGTGTGG	2214
Qy	2507	ataagccctgagtaggcaagagcagtgagatccaactgctatggtcttgatacatccctca	2566
Db	2215	ATGAAGCCTTGAGCCACAACAGGACAGCGGATCCACAGTCATGGTTTCCATGCACCCTC-	2273
Qy	2567	aacttccctcccagcacagaggaattatgctggctgcacacctgcaaaagaaa	2622
Db	2274	TCCCTTCCCTCCCGACACACTGGAGTATTGCCCTGGCATGTAAACCTGCAAAAGAAA	2329
RESULT	8		
AK025880			
LOCUS		Homo sapiens CDNA: FLJ22227 fis, clone HRC01782, highly similar to	
DEFINITION		HS387747 Homo sapiens mRNA for sialin.	
ACCESSION		AK025880	
VERSION		AK025880.1 GI:10438530	
KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Homo sapiens primary human renal epithelial cells cdna to mrna, clone_lib:HRC clone:HRC01782.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		1 (sites) Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S. NEDO human cDNA sequencing project Unpublished (2000)	
TITLE		2 (bases 1 to 1587)	
JOURNAL		Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y. Direct Submission	
REFERENCE		Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ins-u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)	
AUTHORS		NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
FEATURES		Location/Qualifiers	
source		1..1587 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_type="primary human renal epithelial cells" /clone="HRC01782" /clone.lib="HRC" /note="cloning vector pME18SFL3"	
misc_feature		1..1587 /note="highly similar to HSA387747 Homo sapiens mRNA for sialin"	
BASE COUNT		427 a 349 c 350 g 461 t	
ORIGIN			
Query Match		53.4%; Score 1565.8; DB 9; Length 1587;	
Best Local Similarity		99.9%; Pred. No. 0;	
Matches 1567; Conservative		0; Mismatches 2; Indels 0; Gaps 0;	
Qy	1362	agaatgggttttatcttcattgccttatttagctcttgttatgatgatcctgctgt	1421
Db	18	AGAATGGGTTTTATCTTCATGTGCTTATTTAGCTCTTGTTAATGATGATGCTGCTG	77
Oy	1422	gtcaagctgctgcacaatttaaaggcacaatggaatttttcaactttatgtgttcgcagaa	1481

Db 78 GTCAAGCTGCTGACAAATTTAAAGGGCAAAATGGAATTTTCAACTTTATGTGTCGAGAA 137
|||||
QY 1482 tttttagccttagaagattgagcctgcagtgatctcctgtagtgcctgcttcattg 1541
|||||
Db 138 TTTTTFAGCCCTTAGAATGATGACCGTCAGATATCTCGTAGCTGCTGCTTCATTG 197
|||||
QY 1542 gctgtagattattctttggccgttgctttcttaactatatcaacaacactggaggctttt 1601
|||||
Db 198 GCTGTGATTATTTCTTTGGCCGTGCTTTCTTAACATATATCAACAACACTGGGAGGCTTTT 257
|||||
QY 1602 gctcttcgatttagcatcaaccatctgatatctgctccttcgtagtgcgttatcctcc 1661
|||||
Db 258 GCTCTCTCGATTAGCATCAACCATCTGGATATTGCTCCTTCGTATGCTGTATCTCTCC 317
|||||
QY 1662 tgggcatcaacaatacatttgcactattccagggaatggttgggcccgcattgctcaaaa 1721
|||||
Db 318 TGGGCATCACAAATACATTTGCCACTATTCCAGGAATGTTGGGCCCGTCATTGCTTAAA 377
|||||
QY 1722 gctgacccctgataacactgttgagaatggcaaacctgttctctatatattgctgcta 1781
|||||
Db 378 GTCTGACCCCTGATAACACTGTTGAGAAATGGCAAAACCGTGTCTATATTGCTGTGCTA 437
|||||
QY 1782 ttaatgttttggccattttctttacactattcgcaaaagtggaagtacaaaactggg 1841
|||||
Db 438 TTAATGTGTTTTGGTGCCATTTTCTTTACACTATTTCGCCAAAGGTGAAGTACAAAACCTGG 497
|||||
QY 1842 ctctcaatgatcaccatggacacacactgaaggaaccaataaaatcctgcctctat 1901
|||||
Db 498 CTCTCAATGATCACCATGGACACACACTGAAGGACCATAAATAATCCCTGCCCTCAT 557
|||||
QY 1902 taatgtatttttattatcatgtaacctcaagtgctctctgtattgtgtaagcatctcta 1961
|||||
Db 558 TAATGTATTTTATTTATCATGTAACCTCAAAAGTGCTTCTGTATTGTGTAAGCATCTA 617
|||||
QY 1962 tctcttttttaattgactctattagatttttaaggcctataatcatgaataatcaact 2021
|||||
Db 618 TGTCTTTTTTAAATGTACTGTATTAGATTTTAAAGGCTCTATAATCATGAATAATPACT 677
|||||
QY 2022 agttgccagaataaaaaatgaactgtgtttaaattatgaataatgtaagcttagact 2081
|||||
Db 678 AGTTGCCAGAAATAAATAACACTGTGTGTTAATTATGAATAATATGTAAGCTAGACTT 737
|||||
QY 2082 ctacttttaggttcaatacactgcctgctagctgtagtcggcaacatgaagttagacagttctgt 2141
|||||
Db 738 CTACTTTTAGGTTTCATATACCTGCCCTGCTAGTCGGCAACATGAAGTAGGACAGTTCTGTT 797
|||||
QY 2142 gatttttagggccataactaaagggaatgagctgaaacagacctcctgatacctttgctt 2201
|||||
Db 798 GATTTTTTAGGGCCACTAAAGGGAATGAGCTGAACAGACCTCCTGTATACCTTTGCTT 857
|||||
QY 2202 aataaactagatgataattctcagtgactgataaacacactgttgttgcactttcctc 2261
|||||
Db 858 AATTAACCTAGATGATAATCTCAGGTACTGATAAACACTGTGTGTGTCATTTCCTC 917
|||||
QY 2262 ataaaaattgtagctctctctgacacttagacctcaaaacttagatcctctgtgagact 2321
|||||
Db 918 ATAAAAATTGTGAGCCCTCTCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGAGCT 977
|||||
QY 2322 gccatccactgataatttcgctggcaactggaactgaaggagtggtgcccaggagctg 2381
|||||
Db 978 GCATCCACTGTATTAATTTGCGCTTGCAACTGGACTGAGGGGAGTGTGCCAGGACGCTG 1037
|||||
QY 2382 ccaagcaactccctccctgcttcagggtcagagtgcgccagcgttttatcagagagcagcatc 2441
|||||
Db 1038 CCAAGCACTCCCTCCCTGGCTTCAGGTCAGAGTGCCAGAGCTTTATCAGAGCAGCATC 1097
|||||
QY 2442 caagcccagagccagtgctgactcttcggctggtgcctttcctctgaggggctatactg 2501
|||||
Db 1098 CAAGCCAGAGCCAGTGTGCACTCTTCGGCTGCTGCTTCTCTCTGAGGGCTATCAATG 1157
|||||
QY 2502 ttgatataaagccctgagtaggaagagcagtgagatccaactgcttatggtctctgtatacat 2561
|||||

Db 1158 TGTAGATAAAGCCCTGAGTAGGCAAGAGCAGTGAGATCCACTGCTATGTGTGTATACAT 1217
QY 2562 cctcaaacctttccctccagcacagagaaattattgctgggaatgcaactcgcaaaagaa 2621
|||||
Db 1218 CCTCAAACTTTCCCTTCCAGCACACAGAGGAATATTGGCTGGCATGCAACTGCAAAAGAA 1277
|||||
QY 2622 aaatgcgaagcggcggggcacggtggtcctcatgctctgtaatccccagcactttggggggctg 2681
|||||
Db 1278 AAATGCGAAAGCGCGCGCACGGTGGCTCATGCTGTAATCCAGCACTTTGGGGGGCTG 1337
|||||
QY 2682 agttggggcaatcatgagatcagaggagttcagacacagcctggccagcatggtgaaacccc 2741
|||||
Db 1338 AGTGGGCGGAATCATGAGATCAGAGTTTCAGACCAGCTGGCCAGCATGGTGAACCCCC 1397
|||||
QY 2742 atctctactaaaaatacaaaaattagctggcgctggtgacggcgccctgtaatccacaga 2801
|||||
Db 1398 ATCTCTACTTAAATAACAAAAAATTAGCTGGGCTGTGTCACGGCGCCTGTATATCCAGA 1457
|||||
QY 2802 tactcaggggctgaggttaggagaatcacttgaacctggaggtggaagtgcagtgaaac 2861
|||||
Db 1458 TACTCAGGAGGCTGAGGTAGGAGAAATCACTTGAACCTGGGAGGTGGAAGTTGCAGTGAAC 1517
|||||
QY 2862 caagatcacgccactcactccagcctggcgatggcagagagactccaactcaaaaaaa 2921
|||||
Db 1518 CAAGATCACGCCACTGCATCTCCAGCCTGGGCGATGGCGAGACTCCAACCTCAAAAAAA 1577
|||||
QY 2922 aaaaaaaaaa 2930
|||||
Db 1578 AAAAAAAA 1586
|||||

RESULT 9
AK026921
LOCUS
DEFINITION Homo sapiens cDNA: FLJ23268 fis, clone COL08932, highly similar to HSA387747 Homo sapiens mRNA for slalin.
ACCESSION AK026921
VERSION AK026921.1 GI:10439893
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL08932.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2006)
2 (bases 1 to 2006)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES
Location/Qualifiers
1..2006
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL08932"
/clone_lib="COL"
/tissue_type="colon"

/note="cloning vector pME18SFL3"																							
1..2006																							
/note="highly similar to HSA387747 Homo sapiens mRNA for sialin"																							
misc_feature	559 a	418 c	420 g	609 t																			
BASE COUNT																							
ORIGIN																							
Query Match 45.6%; Score 1336.4; DB 9; Length 2006;																							
Best Local Similarity 99.9%; Pred. No. 3.2e-305;																							
Matches 1337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;																							
QY	1593	gagcgctttgctctctggtatttagcatcaaccatctggattgtccctctcgtatgctg	1652																				
Db	1	GAGGCTTTTGCTCTCTCGGATTTAGCATCAACCATCTGGATATTGCTTCTTCGTATGCTG	60																				
QY	1653	gtatcctctctgggcatacacaaatacatttgccactattccaggaaatggttggggcccgctca	1712																				
Db	61	GTATCCTCTCTGGGCATCACAAATACATTTGCCACTATTCCAGGAATGGTTGGGCGCCGTCA	120																				
QY	1713	ttgctaaagctcgacccctgataaacactgttgagaatggcaaacggtgttctctattg	1772																				
Db	121	TTGCTAAAAGTCTGACCCCTGATAAACACTGTTGAGAGAAATGCCAAACCGTGTCTATATTG	180																				
QY	1773	ctgctgctattaatgttttttggcgccattttcttacactattcgcaaaaggtgaagtac	1832																				
Db	181	CTGCTGCTAATTAATGTTTTGGTGCCATTTTCTTTACACTATTTCGCCAAAGGTCAAGTAC	240																				
QY	1833	aaaactgggctctcaatgatcaccatggacacagacactgaaggaacccaataataatcc	1892																				
Db	241	AAAACTGGGCTCTCAATGATCACCATGGACACAGACACTGAAGGAACCAATAATAATCC	300																				
QY	1893	tgctctattaatgtattttattatcatgttaacctcaaatgacctctctctgtattgtga	1952																				
Db	301	TGCTCTAATTAATGATTTTATTATTCATGTAACCTCAAAAGTCCCTTCGTATTGTGTA	360																				
QY	1953	agcattctatgtcttttttaattgtacttattagatttttaagccctataatcatga	2012																				
Db	361	AGCATTCATGCTCTTTTTTAATGTACTTGTATTAGATTTTTTAAGGCCCTAATAATCATGA	420																				
QY	2013	aatacactagtgtccagaataataaaatgaactgtgtttaattatgaataatgtgaag	2072																				
Db	421	AATATCACTAGTTGCCAGAAATAATAAATGAACCTGTGTTTAATTAATGAATAATATGAAG	480																				
QY	2073	ctaggacttctacttttaggttccacatcactgcctgctagtcgggcaacatgaagttagac	2132																				
Db	481	CTAGGACTTCTACTTTAGGTTTCATACCTGCTGCTAGTCGGGCAACATGAAGTAGGAC	540																				
QY	2133	agttctgtgattttttagggccatactaaagggaatgagctgaaacagacctcctgata	2192																				
Db	541	AGTTCTGTGATTTTTTAGGCCATACTAAAGGGAATGAGCTGAAACAGACCTCCTCATGA	600																				
QY	2193	cttttgccttaattaaactagatgataattctcaggtaactgataaaacacctgttgttttc	2252																				
Db	601	CCTTTGCTTAATTAAACTAGATGATAAATCTCAGGTACTGATAAACCACCTGTTGTTGTTTC	660																				
QY	2253	actttctcataaaaattgtcgactctctctgacacttagacctcaaaactttagcatctc	2312																				
Db	661	ACTTTCTCATAAAAATTGTCAGCTCTCTCTGACACTTAGACCTCAAACTTTAGCATCTC	720																				
QY	2313	tgtggagctgccatccactgtataatttcgcctggcacaactggagctgagggagtggtccc	2372																				
Db	721	TGTGGAGCTGCCATCCACTGTATAAATTCGCCCTGGCAACTGGACTGAGGGAGTGTGCC	780																				
QY	2373	aggcagctccaagcactccctccctcgtctcagggttcagagtgccagcggtttatcaga	2432																				
Db	781	AGGCAGCTGCCAAGCACTCCCTCCCTGGCTTCAGGGTTCAGAGTGCAGAGTGCACAGCGTTTATCAGA	840																				
QY	2433	ggcagcatccaagccagagccagtgctgactcttgcctgggtggccttctctctcgtgaggg	2492																				
Db	841	GGCAGCATCCAAGCCCCAGCCAGTGTGCACTCTTCGGCTGGTGCCTTTCTCTGAGGGG	900																				
QY	2493	ctatcaatgtgtagataaaagccctgagttaggcaagagcagctgagatccactgctatgctc	2552																				

Db	901	CTATCAATGTGTAGATAAAGCCCTGAGTAGGCAAGACGACGTGAGATCCACTGCTATGGTC	960																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
----	-----	---	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

56641 bp DNA linear HTG 10-JUL-2001

56641 bp DNA linear HTG 10-JUL-2001

56641 bp DNA linear HTG 10-JUL-2001

56641 bp DNA linear HTG 10-JUL-2001

56641 bp DNA linear HTG 10-JUL-2001

56641 bp DNA linear HTG 10-JUL-2001

* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 35016: contig of 35016 bp in length
* 35017 35116: gap of 100 bp
* 35117 38963: contig of 3847 bp in length
* 38964 39063: gap of 100 bp
* 39064 56641: contig of 17578 bp in length.

FEATURES

source

Location/Qualifiers
1..56641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP3-515P18"
/clone_lib="RPCI-3"
1..35016
/note="assembly_fragment:00367
clone_end:n7
vector_side:left"
35117..38963
/note="assembly_fragment:00264
fragment_chain:1"
39064..56641
/note="assembly_fragment:00607
fragment_chain:1"

misc_feature

misc_feature

misc_feature

BASE COUNT 16181 a 12335 c 12027 g 15895 t 203 others
ORIGIN

Query Match 40.7%; Score 1193.6; DB 2; Length 56641;
Best Local Similarity 99.3%; Pred. No. 3.le-271;
Matches 1199; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1722 gctgacccctgataacacgtgtgagaaatggcaaacctgttctatatgtctgtgcta 1781
|||||
DB 43034 GTCCTCTTTATAGAACACTGTTGGAGATGCGCAACCGTGTCTATATATGCTGCTA 43093
|||||
QY 1782 ttaatgtttttggtgccattttcttacactattcgccaaagtggaagtacaaactggg 1841
|||||
DB 43094 TTAATGTTTTTGTGGCCATTTCTTTACACTATTGCCAAAGTGAAATACAAAACCTGGG 43153
|||||
QY 1842 ctctcaatgatacaccatgagacacacactgaaggaaccaataataatcctgccttat 1901
|||||
DB 43154 CTCTCAATGATCACCATGACACAGACACTGAAGGAACCAATAATAATCCTGCCCTTAT 43213
|||||
QY 1902 taatgtattttatttatcatgtaacctcaaatgacctgtctgtattgtgaagcattcta 1961
|||||
DB 43214 TAATGTATTTTATTATCATGTAACCTCAAGTGCCCTCTGTATTGTGTAAGCATTTCTA 43273
|||||
QY 1962 tgccttttttaattgtactgtattagatttttaaggccttaataatcatgaataatcact 2021
|||||
DB 43274 TGCTTTTTTTAATGTACTGTATTAGATTTTAAAGGCTATAATCATGAAATATCACT 43333
|||||
QY 2022 agttgccagaataataaaatgaactgttttaattatgaataatgtaagcttagactt 2081
|||||
DB 43334 AGTTGCCAGAAATAAATAAGTAACCTGTGTTAATTATGAATAATATGTAAGCTAGACTT 43393
|||||
QY 2082 ctacttttagttcacatacctgcctgctagtgcgggcaacatgaagttagacagttctgtt 2141
|||||
DB 43394 CTACTTTAGTTTCATACATCCTGCTAGCTAGTCGGGCAACATGAAGTAGGACAGTTCTGT 43453
|||||
QY 2142 gatcttttagggccatactaaagggaatgagctgaacagacctcctgatacctttgctt 2201
|||||
DB 43454 GATTTTTTAGGGCCATCTATAAGGGAATGAGCTGAACAGACCTCCTGTACTCTTGTCTT 43513
|||||
QY 2202 aattaaactagatgataatctcaggtactgataaaacacctgttgttgcactttcctc 2261
|||||
DB 43514 AATTAACCTAGATGATAATCTCAGGTACTGATGATAAACACCTGTGTGTGTTCACTTTCC 43573
|||||

QY 2262 ataaaaattgtcagctctctctgacaccttagacacctcaaaacttttagcatctctgtgagct 2321
|||||
DB 43574 ATAAAAATTGTGACGCTCTCTCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGAGCT 43633
|||||
QY 2322 gccatccactgtataaatttcgctcctggaactggaactgaggtgaggtgccccagggcagctg 2381
|||||
DB 43634 GGCATCCACTGTATATATTTCGCTTGGCTGGCAACTGAGCTGAGGGAGTGTGCCAGGACGCTG 43693
|||||
QY 2382 caaagcaactccctcctggttcagggtcagagtcagagtcgcccagcttttataagagcaacatc 2441
|||||
DB 43694 CAAAGCACTCCCTCCCTGGCTTCAGGGTCAGAGTGCCCCAGCGTTTATCAGAGGACGATC 43753
|||||
QY 2442 caagcccagagccaggtgtcgaactcttcggctggtgcttctcctcctgaggggctatcaatg 2501
|||||
DB 43754 CAAGCCACAGGACGAGTGTGCACTCTTCGGCTGGTGGCTTTCTCTGAGGGCTATCAATG 43813
|||||
QY 2502 tgtagataaagccctgagtaggcaagagcagtgagatccactgctatgcttcttgcatacat 2561
|||||
DB 43814 TGTAGATAAAGCCCTTGAGTAGGCAGAGAGAGTGCAGATCCACTGCTATGCTTGTATACAT 43873
|||||
QY 2562 cctcaaaactttccctcccccagcacagagaataattgctggcagcatgcaacctgcaaaagaa 2621
|||||
DB 43874 CCTCAAACTTTCCCTTCCCGAGCAGAGGAATATTGGCTGGCATGCAACTGCAAAAGAA 43933
|||||
QY 2622 aaatgcgaagcgccgggcacggtggtcctatgctcctgtaatcccgagcacttttgggggctg 2681
|||||
DB 43934 AAATGCGAAGCGCGCGGCGACGGTGGCTCATGCTGTATATCCAGCACTTTGGGGGGCTG 43993
|||||
QY 2682 aggtgggcgaatcatgagatcagggaggttcgagaccagcctggccagcatggttgaaccccc 2741
|||||
DB 43994 AGGTGGCGCAATCATGAGATCAGGAGTTGCGAGACCAGCCTGGCCAGCATGGTGAACCC 44053
|||||
QY 2742 atctctactaaaaatacaaaaaattagctgggctgtgtgacggcgccctgtaatccaca 2801
|||||
DB 44054 ATCTCTACATAAAATAACAAAAATTAGCTGGCGGTGGTGACGGGCGCTGTATATCCAGA 44113
|||||
QY 2802 tactcaggagctgaggttaggagaaatcactgaaacctggaggtggaagtgcagtgagac 2861
|||||
DB 44114 TACTCAGGAGCTGAGGTAGGAGAACTTGAACCTGGAGGTGGAAGTTGCAGTGAAC 44173
|||||
QY 2862 caagatcacgcacactcactccagcctggcgatggagcgagactcccaactcaaaaaaa 2921
|||||
DB 44174 CAAGATCACGCCACTGCATCCAGCCTGGCGCATGGAGCGAGACTCCAACCTCAAAAAAAA 44233
|||||
QY 2922 aaaaaaaa 2929
|||||
DB 44234 AAAAGAA 44241
|||||

RESULT 11

HSJ397H23/c

LOCUS

Human DNA sequence from clone RP3-397H23 on chromosome 6q12-14.1,
complete sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

DIRECT SUBMISSION

SUBMITTED (17-MAR-2001)

CBIO 15A, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Mar 12, 2001 this sequence version replaced gi:12832031.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP3-397H23 is from the library RP3-3 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP3-397H23.

FEATURES

source	Location/Qualifiers
	1..113202
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/map="q12-14.1"
	/clone="RP3-397H23"
	/clone_lib="RP3-3"
	1..58
repeat_region	/note="AluJ/FLAM repeat: matches 5..62 of consensus"
repeat_region	260..570
repeat_region	/note="AluSx repeat: matches 1..305 of consensus"
repeat_region	674..707
repeat_region	/note="MER20 repeat: matches 29..62 of consensus"
repeat_region	708..880
repeat_region	/note="AluSp repeat: matches 120..292 of consensus"
repeat_region	961..1095
repeat_region	/note="AluSg1 repeat: matches 1..135 of consensus"
repeat_region	1105..1428
repeat_region	/note="AluSx repeat: matches 1..306 of consensus"
repeat_region	1432..1585
repeat_region	/note="MER20 repeat: matches 60..218 of consensus"
repeat_region	1826..2123
misc_feature	/note="AluY repeat: matches 1..295 of consensus"
misc_feature	complement(2112..2118)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data"
misc_feature	complement(2120..2127)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data"
misc_feature	complement(2129..2147)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data"
misc_feature	complement(2149..2160)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data"
misc_feature	complement(2168..2263)
repeat_region	2150..2458
misc_feature	/note="AluSx repeat: matches 1..306 of consensus"
misc_feature	complement(2161..2164)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data"
misc_feature	complement(2168..2263)
misc_feature	/note="Single clone region. assembly confirmed by restriction digest data"
misc_feature	complement(2168..2263)
repeat_region	2468..2774
repeat_region	/note="AluSg1 repeat: matches 1..308 of consensus"
repeat_region	2795..2912
repeat_region	/note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region	3028..3325
repeat_region	/note="AluY repeat: matches 1..296 of consensus"
repeat_region	3355..3644
repeat_region	/note="AluSg repeat: matches 1..287 of consensus"
repeat_region	3775..4074
repeat_region	/note="AluSx repeat: matches 1..300 of consensus"
repeat_region	4096..4386
repeat_region	/note="AluJo repeat: matches 1..298 of consensus"
repeat_region	4490..4793
repeat_region	/note="AluJb repeat: matches 1..306 of consensus"
repeat_region	5117..5253
repeat_region	/note="FLAM_C repeat: matches 1..132 of consensus"
repeat_region	5501..5814
repeat_region	/note="AluSg repeat: matches 1..308 of consensus"
repeat_region	6097..6225
repeat_region	/note="FLAM_A repeat: matches 1..131 of consensus"
repeat_region	6296..6479
repeat_region	/note="AluJb repeat: matches 129..312 of consensus"
repeat_region	6545..6830
repeat_region	/note="AluJo repeat: matches 42..306 of consensus"
repeat_region	6965..7270
repeat_region	/note="AluSg repeat: matches 1..306 of consensus"
repeat_region	7308..7602
repeat_region	/note="AluJb repeat: matches 1..295 of consensus"
repeat_region	7737..8035
repeat_region	/note="AluSg repeat: matches 1..299 of consensus"
repeat_region	8184..8488
repeat_region	/note="AluSx repeat: matches 1..299 of consensus"
repeat_region	8917..9176
repeat_region	/note="AluSg1 repeat: matches 1..293 of consensus"
repeat_region	9212..9377
repeat_region	/note="L1MB4 repeat: matches 6005..6177 of consensus"
repeat_region	9490..9785
repeat_region	/note="AluSx repeat: matches 1..296 of consensus"
repeat_region	10352..10431
repeat_region	/note="L2 repeat: matches 2616..2702 of consensus"
repeat_region	10432..10734
repeat_region	/note="AluSx repeat: matches 1..301 of consensus"
repeat_region	10735..10778
repeat_region	/note="L2 repeat: matches 2575..2616 of consensus"
repeat_region	10846..10985
repeat_region	/note="MIR repeat: matches 53..200 of consensus"
repeat_region	11023..11140
repeat_region	/note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region	12005..12079
repeat_region	/note="L1MD3 repeat: matches 7662..7737 of consensus"
repeat_region	12350..12647
repeat_region	/note="AluSx repeat: matches 1..297 of consensus"
repeat_region	12669..12916
repeat_region	/note="AluSx repeat: matches 1..293 of consensus"
repeat_region	13058..13351
repeat_region	/note="AluSc repeat: matches 1..296 of consensus"
repeat_region	13504..13657
repeat_region	/note="AluY repeat: matches 152..306 of consensus"
repeat_region	13838..13947
repeat_region	/note="AluSx repeat: matches 25..138 of consensus"
repeat_region	13948..14254
repeat_region	/note="AluSg repeat: matches 1..305 of consensus"
repeat_region	14255..14437
repeat_region	/note="AluSx repeat: matches 1..301 of consensus"
repeat_region	14934..15229
repeat_region	/note="AluSp repeat: matches 1..297 of consensus"
repeat_region	15266..15556
repeat_region	/note="AluJo repeat: matches 5..299 of consensus"
repeat_region	15578..15888
repeat_region	/note="AluYa5 repeat: matches 1..311 of consensus"
repeat_region	15965..16082
repeat_region	/note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region	16157..16439
repeat_region	/note="AluJb repeat: matches 11..294 of consensus"
repeat_region	16618..16913
repeat_region	/note="AluSp repeat: matches 1..296 of consensus"

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lien,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEvan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meidirim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 12, 2000 this sequence version replaced gi:7417823.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7098
 Center Clone name: 598_G_19
 ----- Summary Statistics

TITLE
 JOURNAL
 COMMENT

Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 134701 bases at least Q40
 Consensus quality: 141812 bases at least Q30
 Consensus quality: 144823 bases at least Q20
 Insert size: 166000; agarose-gel
 Quality coverage: 3.6 in Q20 bases; sum-of-contigs
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 1018: contig of 1018 bp in length
 * 1019 1118: gap of 100 bp
 * 1119 2203: contig of 1085 bp in length
 * 2204 2303: gap of 100 bp
 * 2304 3814: contig of 1511 bp in length
 * 3815 3914: gap of 100 bp
 * 3915 4327: contig of 413 bp in length
 * 4328 4427: gap of 100 bp
 * 4428 6447: contig of 2020 bp in length
 * 6448 6547: gap of 100 bp
 * 6548 8884: contig of 2337 bp in length
 * 8885 8984: gap of 100 bp
 * 8985 10630: contig of 1646 bp in length
 * 10631 10730: gap of 100 bp
 * 10731 13203: contig of 2472 bp in length
 * 13203 13302: gap of 100 bp
 * 13303 15992: contig of 2690 bp in length
 * 15993 16092: gap of 100 bp
 * 16093 18722: contig of 2630 bp in length
 * 18723 18822: gap of 100 bp
 * 18823 21695: contig of 2873 bp in length
 * 21696 21795: gap of 100 bp
 * 21796 25265: contig of 3470 bp in length
 * 25266 25365: gap of 100 bp

25366 28014: contig of 2649 bp in length
 * 28015 28114: gap of 100 bp
 * 28115 31378: contig of 3264 bp in length
 * 31379 31478: gap of 100 bp
 * 31479 34284: contig of 2806 bp in length
 * 34285 34384: gap of 100 bp
 * 34385 37707: contig of 3323 bp in length
 * 37708 37807: gap of 100 bp
 * 37808 41922: contig of 4115 bp in length
 * 41923 42022: gap of 100 bp
 * 42023 45878: contig of 3856 bp in length
 * 45879 45978: gap of 100 bp
 * 45979 49890: contig of 3912 bp in length
 * 49891 49990: gap of 100 bp
 * 49991 54041: contig of 4051 bp in length
 * 54042 54141: gap of 100 bp
 * 54142 59531: contig of 5390 bp in length
 * 59532 59631: gap of 100 bp
 * 59632 65537: contig of 5906 bp in length
 * 65538 65637: gap of 100 bp
 * 65638 71486: contig of 5849 bp in length
 * 71487 71586: gap of 100 bp
 * 71587 76694: contig of 5108 bp in length
 * 76695 76794: gap of 100 bp
 * 76795 82579: contig of 5785 bp in length
 * 82580 82679: gap of 100 bp
 * 82680 90920: contig of 8241 bp in length
 * 90921 91020: gap of 100 bp
 * 91021 99034: contig of 8014 bp in length
 * 99035 99134: gap of 100 bp
 * 99135 110324: contig of 11190 bp in length
 * 110325 110424: gap of 100 bp
 * 110425 119810: contig of 9386 bp in length
 * 119811 119910: gap of 100 bp
 * 119911 130142: contig of 10232 bp in length
 * 130143 130242: gap of 100 bp
 * 130243 149597: contig of 19355 bp in length.

FEATURES
 Location/Qualifiers

1..149597
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="6"
 /clone="RP11-598G19"
 /clone_lib="RPC1-11 Human Male BAC"
 1..1018
 /note="assembly_fragment"
 misc_feature
 1119..2203
 /note="assembly_fragment"
 misc_feature
 2304..3814
 /note="assembly_fragment"
 misc_feature
 3915..4327
 /note="assembly_fragment"
 misc_feature
 clone_end:r7
 vector_side:right
 4428..6447
 /note="assembly_fragment"
 misc_feature
 6548..8884
 /note="assembly_fragment"
 misc_feature
 8985..10630
 /note="assembly_fragment"
 misc_feature
 10731..13202
 /note="assembly_fragment"
 misc_feature
 13303..15992
 /note="assembly_fragment"
 misc_feature
 16093..18722
 /note="assembly_fragment"
 misc_feature
 18823..21695
 /note="assembly_fragment"
 misc_feature
 21796..25265
 /note="assembly_fragment"
 misc_feature
 25366..28014
 /note="assembly_fragment"

```
misc_feature 28115..31378
/note="assembly_fragment"
misc_feature 31479..34284
/note="assembly_fragment"
misc_feature 34385..37707
/note="assembly_fragment"
misc_feature 37808..41922
/note="assembly_fragment"
misc_feature 42023..45878
/note="assembly_fragment"
misc_feature 45979..49890
/note="assembly_fragment"
misc_feature 49991..54041
/note="assembly_fragment"
misc_feature 54142..59531
/note="assembly_fragment"
misc_feature 59632..65537
/note="assembly_fragment"
misc_feature 65638..71486
/note="assembly_fragment"
misc_feature 71587..76694
/note="assembly_fragment"
misc_feature 76795..82579
/note="assembly_fragment"
misc_feature 82680..90920
/note="assembly_fragment"
misc_feature 91021..99034
/note="assembly_fragment"
misc_feature 99135..110324
/note="assembly_fragment"
misc_feature 110425..119810
/note="assembly_fragment"

Query Match 40.7%; Score 1192; DB 2; Length 149597;
Best Local Similarity 99.2%; Pred. No. 8.9e-271;
Matches 1198; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1722 gtctgacccctgatacacactgttgagaaatggcaaacctgttctctattattgctgctgcta 1781
|||||
DB 75367 GTCTTCTCTATAGAACACTGTTGGAGATGGCAACCGTTCATATTATGCTGCTGCTA 75426

QY 1782 ttaatgttttggcgcaattttcttacactattcgccaaaggatgaagtcacaaactggg 1841
|||||
DB 75427 TTAATGTTTTTGGTGGCATTTCTTTACACTATTTCGCAAGGTGAAGTACAAACCTGGG 75486

QY 1842 ctctcaatgacaccatgacacagacactgaaggaaaccaataaataactctgcctctat 1901
|||||
DB 75487 CTCTCAATGATCACCATGGACACAGACACTGAAGGAACCAATAAATAATCCTGCCCTCTAT 75546

QY 1902 taatgtatttttatcatgttaacctcaaaagtgcctctctgtattgtgttaagcattcta 1961
|||||
DB 75547 TAATGTATTTTATTTATCATGTAACTCAAAAGTCTTCTGTATTGTGTAAAGCATTTCTA 75606

QY 1962 tgccttttttaattgtactgtattagatttttaaggccctataatgaataatacaact 2021
|||||
DB 75607 TGTCFTTTTTTAATTGCTACTGTATTAGATTTTAAAGGCCATATAAATCATGAAATATCACT 75666

QY 2022 agttgcagaataataaatagaactgttttaattatgaataataatgaactagacatt 2081
|||||
DB 75667 AGTTGCCAGAATAAATAATGAACCTGTGTTAATTATGAATAATATGTAAGCTAGGACTT 75726

QY 2082 ctacttttagttcacatacctgcctgctagtcgggcaacatgaagttaggacagttctgtt 2141
|||||
DB 75727 CTACTTTAGTTTCACATACCTGCTGCTAGTCGGGCAACATGAAGTAGGACACTTCTGTT 75786

QY 2142 gatcttttggggccatactaaaggaatgagctgaacagacctctgtatcaccttgcctt 2201
|||||
DB 75787 GATTTTTTAGGGCCACTAAAGGAATGAGCTGGAACACAGACCTCTCTGATACCTTTGCTT 75846

QY 2202 aattaactagatgaataattctcagttactgataacacacctgtgtgttcttcaacttctc 2261
|||||
DB 75847 AATTAACATAGATGAATAATCTCAGGTACTGATAAACACCTGTTGTGTTCCTCTCTCCT 75906
```

```
QY 2262 ataaaattgtcagctctctctgacacttagaacctcaaaacttttagcatctctgtgagct 2321
|||||
DB 75907 ATAAAAATTGTCAGCTCTCTCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGGAGCT 75966

QY 2322 gccatccactgtataatttgcgcctggcaactggactgaggggagtgccacagggcagctg 2381
|||||
DB 75967 GCCATCCACTGTATAATTTCGCCTGGCACTGGAGCTGAGGGGAGTGTGCCACGACGAGCTG 76026

QY 2382 ccaagcactccctccctggcttcagaggtcagagtcacagcgtttatcagagcagcattc 2441
|||||
DB 76027 CCAAGCACTCCCTCCCTGGCTTCAGGTCAGAGTGCCACGCTTTATCAGAGCAGCAGTATC 76086

QY 2442 caagccagagccagtgctgactcttcggctggctgttcctctcctcctgaggggcttatcaatg 2501
|||||
DB 76087 CAAGCCAGAGCCAGTGTGACTCTTCGGCTGCTGCTCTTTCCCTCTGAGGGGCTATCAATG 76146

QY 2502 tgtagataaagccctgagtaggcaagcagtgagatccaactgctatggtctctgatacat 2561
|||||
DB 76147 TGTAGATAAAGCCCTGAGTAGGCAAGCAGCAGTGAATCCACTGCTATGCTTGTATACAT 76206

QY 2562 cctcaaaacttcccttcccagcacagaggaatataggctgagcatgcaacctgcaaaagaa 2621
|||||
DB 76207 CCTCAAACTTTCCCTTCCCAGCACAGAGGAATATTTGCTGGCATGCAACCTGCAAAAGAA 76266

QY 2622 aaatgcaagcgccggcgacggtggctcatgctctgtaatcccagcactttgggggctg 2681
|||||
DB 76267 AAATGCGAAGCGCGCGGACAGGTGGCTCATGCTGTATATCCCAGCACCTTTTGGGGGCTG 76326

QY 2682 agtgggcggaatcatgatcagaggttcagaccagcctggccagcatggtgaaacccc 2741
|||||
DB 76327 AGTGGGCGAATCATGAGATCAGAGTTTCGAGACCAGCCTTGGCCAGCATGGTGAACCCCC 76386

QY 2742 atcttactaaaaatacaaaaaattagctggcgctggtgacggcgccctgtaatacccaga 2801
|||||
DB 76387 ATCTTACTAAAAATACAAAAAATTAGCTGGCGTGTGTGACGGCGCCTGTAAATCCAGAA 76446

QY 2802 tactcaggagcggtgagtagagaatcacttgaacctggaggtggaggttgagtgagtgaaac 2861
|||||
DB 76447 TACTCAGGAGGCTGAGGTAGGAGAATCACATTGAACCTGGGAGATGGAAGTTGCAGTGAAC 76506

QY 2862 caagatacagccactgcactccagcctggcgatgagcagagactccaactcaaaaaaaa 2921
|||||
DB 76507 CAAGATCACCCCACTGCACCTCCAGCCTGGCGCATGGAGCGAGACTCCCACTCAAAAAAAA 76566

QY 2922 aaaaaaaa 2929
|||||
DB 76567 AAAAGAA 76574

RESULT 13
AC025535/c
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-206H23, WORKING DRAFT
SEQUENCE 25 unordered pieces.
AC025535
VERSION AC025535.4 GI:7658462
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157749)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157749)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Apr 28, 2000 this sequence version replaced gi:7582731.
```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0206H23
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator; plasmid; 0%
Chemistry: Dye-terminator; Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 141377 bases at least Q40
Consensus quality: 146682 bases at least Q30
Consensus quality: 148842 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 153549; sum-of-contigs
Quality coverage: 3.52 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

	1	1470:	contig	of 1470	bp in length
*	1471	1570:	gap of	unknown	length
*	1571	3039:	contig of	1449	bp in length
*	3020	3119:	gap of	unknown	length
*	3120	5356:	contig of	2237	bp in length
*	5357	5456:	gap of	unknown	length
*	5457	7747:	contig of	2291	bp in length
*	7748	7847:	gap of	unknown	length
*	7848	9213:	contig of	1366	bp in length
*	9214	9313:	gap of	unknown	length
*	9314	10753:	contig of	1440	bp in length
*	10754	10853:	gap of	unknown	length
*	10854	12243:	contig of	1390	bp in length
*	12244	12343:	gap of	unknown	length
*	12344	14531:	contig of	2208	bp in length
*	14532	14651:	gap of	unknown	length
*	14652	17086:	contig of	2415	bp in length
*	17087	17166:	gap of	unknown	length
*	17167	19602:	contig of	2436	bp in length
*	19603	19702:	gap of	unknown	length
*	19703	23843:	contig of	4141	bp in length
*	23844	23943:	gap of	unknown	length
*	23944	27763:	contig of	3840	bp in length
*	27764	27883:	gap of	unknown	length
*	27884	31792:	contig of	3909	bp in length
*	31793	31892:	gap of	unknown	length
*	31893	35687:	contig of	3795	bp in length
*	35688	35787:	gap of	unknown	length
*	35788	39711:	contig of	3924	bp in length
*	39712	39811:	gap of	unknown	length
*	39812	44204:	contig of	4393	bp in length
*	44205	44304:	gap of	unknown	length
*	44305	48833:	contig of	4529	bp in length
*	48834	48933:	gap of	unknown	length
*	48934	56617:	contig of	7684	bp in length
*	56618	56717:	gap of	unknown	length
*	56718	63063:	contig of	6346	bp in length
*	63064	63163:	gap of	unknown	length
*	63164	74288:	contig of	11125	bp in length
*	74289	74388:	gap of	unknown	length
*	74389	87403:	contig of	13015	bp in length
*	87404	87503:	gap of	unknown	length
*	87504	99698:	contig of	12195	bp in length
*	99699	99798:	gap of	unknown	length
*	99799	115384:	contig of	15586	bp in length

[illegible]

QY 2562 cctcaacttccctccagcagaggaattattgctggcatgcaacctgcaaaagaa 2621
|||||
Db 92454 CCTCAACTTTCCTCCAGCAGACAGAGGAATATTGGCTGGCATGCAACCTGCAAAAGAA 92395
QY 2622 aaatgcgaagcggccggcagcgtggtctcatgcttaataccccagcacttggggggctg 2681
|||||
Db 92394 AAATGGGAAGCGCGCGGACGCGTGGCTATGCTCTAATCCAGCAGCACTTTGGGGGGCTG 92335
QY 2682 agtgggcaaatcatagatcaagagttcgagacagcctggccagcatggttaaaacccc 2741
|||||
Db 92334 AGTGGCGCAANTCATGAGATCAGAGTTCGAGACCAGCCTGGCCAGCATGGTGAAACCCC 92275
QY 2742 atctctactaaaaatacaaaaaattagctggcgtggtgacggcgccctgtaattccaga 2801
|||||
Db 92274 ATCTCTACTAAAAATACAAAAATTAGCTGGCGCTGTGACGGCGCGCTGTATATCCAGA 92215
QY 2802 tactcagggagctgaggttaggaaatacaacttgaaactggagctggaggtgcaagtgaac 2861
|||||
Db 92214 TACTCAGGAGGCTGAGGTAGGAGAAATCACTTGAACCTGGGAGATGGAAGTTGCAGTGAAC 92155
QY 2862 caagatcacgccactgcactccagcctggcgcatgagcgagatccaactcaaaaaaa 2921
|||||
Db 92154 CAGATCACGCCACTGCATCTCCAGCTGGCGGATGGAGGAGAGACTCCAACTCAAAAAA 92095
QY 2922 aaaaaaa 2929
|||||
Db 92094 AAAAAGAA 92087

RESULT 14
AL590428/c
LOCUS
DEFINITION Human DNA sequence from clone Rp11-553A21 on chromosome 6, complete sequence.
ACCESSION AL590428 AC028605
VERSION AL590428.7 GI:15072593
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163577)
Direct Submission
Submitted (31-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Aug 1, 2001 this sequence version replaced gi:15021177.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
Rp11-553A21 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
Rp11-553A21. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.

The true right end of clone Rp11-553A21 is at 163577 in this
sequence. The true left end of clone Rp11-525G3 is at 88067 in this
sequence. The true right end of clone Rp3-397H23 is at 2000 in this
sequence.

FEATURES

Source	Location/Qualifiers
1..163577	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="Rp11-553A21"
	/clone_lib="RPCI-11.2"
285..433	
repeat_region	/note="AluJb repeat: matches 1..119 of consensus"
434..743	
repeat_region	/note="AluY repeat: matches 1..311 of consensus"
744..919	
repeat_region	/note="AluJb repeat: matches 119..305 of consensus"
988..1075	
repeat_region	/note="L2 repeat: matches 2626..2708 of consensus"
1166..1384	
repeat_region	/note="L1MC5 repeat: matches 7317..7536 of consensus"
1385..1690	
repeat_region	/note="AluX repeat: matches 1..306 of consensus"
1691..1759	
repeat_region	/note="L1MC5 repeat: matches 7536..7604 of consensus"
1760..2024	
repeat_region	/note="AluY repeat: matches 35..287 of consensus"
2025..2248	
repeat_region	/note="L1MC5 repeat: matches 7604..7887 of consensus"
2588..2706	
repeat_region	/note="L1MC/D repeat: matches 5389..5514 of consensus"
misc_feature	3686..4489
	/note="CpG island"
	/evidence=not_experimental
repeat_region	3695..3736
	/note="21 copies 2 mer cc 78% conserved"
repeat_region	4691..4996
	/note="AluJo repeat: matches 3..306 of consensus"
repeat_region	5628..5977
	/note="MLT2FB repeat: matches 1..403 of consensus"
repeat_region	5978..6030
	/note="MER5A repeat: matches 44..97 of consensus"
repeat_region	6118..6474
	/note="L1MC3 repeat: matches 7398..7739 of consensus"
repeat_region	6475..6771
	/note="AluY repeat: matches 1..290 of consensus"
repeat_region	6772..7373
	/note="L1MC3 repeat: matches 6740..7398 of consensus"
repeat_region	7491..7764
	/note="AluJo repeat: matches 1..286 of consensus"
repeat_region	7887..8168
	/note="AluSq repeat: matches 1..287 of consensus"
repeat_region	8365..8430
	/note="33 copies 2 mer tc 89% conserved"
repeat_region	8431..8729
	/note="AluSq repeat: matches 1..302 of consensus"
repeat_region	8779..8975
	/note="MER33 repeat: matches 129..323 of consensus"
repeat_region	8976..9292
	/note="AluX repeat: matches 1..311 of consensus"
repeat_region	9293..9415
	/note="MER33 repeat: matches 1..129 of consensus"
repeat_region	9516..9647
	/note="MIR repeat: matches 106..226 of consensus"
repeat_region	9698..9994
	/note="AluSq repeat: matches 1..295 of consensus"
repeat_region	10005..10048
	/note="22 copies 2 mer tt 79% conserved"

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 6, clone RP11-598G19
Unpublished
2 (bases 1 to 149597)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Navlor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7417823.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7098
Center clone name: 598_G_19
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134701 bases at least Q40
Consensus quality: 141812 bases at least Q30
Consensus quality: 144823 bases at least Q20
Insert size: 166000; agarose-fp
Insert size: 146597; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* I 1018: contig of 1018 bp in length
* 1019 1118: gap of 100 bp
* 1119 2203: contig of 1085 bp in length
* 2204 2303: gap of 100 bp
* 2304 3814: contig of 1511 bp in length
* 3815 3914: gap of 100 bp
* 3915 4327: contig of 413 bp in length
* 4328 4427: gap of 100 bp
* 4428 6447: contig of 2020 bp in length
* 6448 6547: gap of 100 bp
* 6548 8984: contig of 2337 bp in length
* 8885 8984: gap of 100 bp
* 8985 10630: contig of 1646 bp in length
* 10631 10730: gap of 100 bp

TITLE
JOURNAL
COMMENT

* 10731 13202: contig of 2472 bp in length
* 13203 13302: gap of 100 bp
* 13303 15992: contig of 2690 bp in length
* 15993 16092: gap of 100 bp
* 16093 18722: contig of 2630 bp in length
* 18723 18822: gap of 100 bp
* 18823 21695: contig of 2873 bp in length
* 21696 21795: gap of 100 bp
* 21796 25265: contig of 3470 bp in length
* 25266 25365: gap of 100 bp
* 25366 28014: contig of 2649 bp in length
* 28015 28114: gap of 100 bp
* 28115 31378: contig of 3264 bp in length
* 31379 31478: gap of 100 bp
* 31479 34284: contig of 2806 bp in length
* 34285 34384: gap of 100 bp
* 34385 37707: contig of 3323 bp in length
* 37708 37807: gap of 100 bp
* 37808 41922: contig of 4115 bp in length
* 41923 42022: gap of 100 bp
* 42023 45878: contig of 3856 bp in length
* 45879 45978: gap of 100 bp
* 45979 49890: contig of 3912 bp in length
* 49891 49990: gap of 100 bp
* 49991 54041: contig of 4051 bp in length
* 54042 54141: gap of 100 bp
* 54142 59531: contig of 5390 bp in length
* 59532 59631: gap of 100 bp
* 59632 65537: contig of 5906 bp in length
* 65538 65637: gap of 100 bp
* 65638 71486: contig of 5849 bp in length
* 71487 71586: gap of 100 bp
* 71587 76694: contig of 5108 bp in length
* 76695 76794: gap of 100 bp
* 76795 82579: contig of 5785 bp in length
* 82580 82679: gap of 100 bp
* 82680 90920: contig of 8241 bp in length
* 90921 91020: gap of 100 bp
* 91021 99034: contig of 8014 bp in length
* 99035 99134: gap of 100 bp
* 99135 110324: contig of 11190 bp in length
* 110325 110424: gap of 100 bp
* 110425 119810: contig of 9386 bp in length
* 119811 119910: gap of 100 bp
* 119911 130142: contig of 10232 bp in length
* 130143 130242: gap of 100 bp
* 130243 149597: contig of 19355 bp in length.
* FEATURES
* Location/Qualifiers
* 1..149597
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="6"
* /map="6"
* /clone="RP11-598G19"
* /clone_lib="RPC1-11 Human Male BAC"
* 1..1018
* /note="assembly_fragment"
* 1119..2203
* /note="assembly_fragment"
* 2304..3814
* /note="assembly_fragment"
* 3915..4327
* /note="assembly_fragment"
* clone_end:n7
* vector_side:right
* 4428..6447
* /note="assembly_fragment"
* 6548..8884
* /note="assembly_fragment"
* 8985..10630
* /note="assembly_fragment"
* 10731..13202
* /note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature	13303..15992	/note="assembly_fragment"	
misc_feature	16093..18722	/note="assembly_fragment"	
misc_feature	18823..21695	/note="assembly_fragment"	
misc_feature	21796..25265	/note="assembly_fragment"	
misc_feature	25366..28014	/note="assembly_fragment"	
misc_feature	28115..31378	/note="assembly_fragment"	
misc_feature	31479..34284	/note="assembly_fragment"	
misc_feature	34385..37707	/note="assembly_fragment"	
misc_feature	37808..41922	/note="assembly_fragment"	
misc_feature	42023..45878	/note="assembly_fragment"	
misc_feature	45979..49890	/note="assembly_fragment"	
misc_feature	49991..54041	/note="assembly_fragment"	
misc_feature	54142..59531	/note="assembly_fragment"	
misc_feature	59632..65537	/note="assembly_fragment"	
misc_feature	65638..71486	/note="assembly_fragment"	
misc_feature	71587..76694	/note="assembly_fragment"	
misc_feature	76795..82579	/note="assembly_fragment"	
misc_feature	82680..90920	/note="assembly_fragment"	
misc_feature	91021..99034	/note="assembly_fragment"	
misc_feature	99135..110324	/note="assembly_fragment"	
misc_feature	110425..119810	/note="assembly_fragment"	
misc_feature	119811..120000	/note="assembly_fragment"	
Query Match 13.5%; Score 396.8; DB 2; Length 149597;			
Best Local Similarity 87.1%; Pred. No. 6.9e-83;			
Matches 474; Conservative 0; Mismatches 7; Indels 63; Gaps 1;			
QY	1	gttcggtcgaagccctcccttaattatgtgcaattcaagtccccactgcccgcgcgcaa	60
DB	102673	GTTCGGTCCGAAGCCCTCCCTTAATTATGTGCAATTCAAGTCCCCACTGCCCGCCGCAA	102614
QY	61	gcccccaactcactcgtcgtgaggggtggccctgcaattacaaaggggtgcagga	120
DB	102613	GCCCCCACCACATCTCTCGCTCGGGCAGGGTGGCCCTGCACCTTTACAAGGGGTGCAGGA	102554
QY	121	gcgggagagcgttcgccgaacacggctcccgcgcatgtagaccgagcgagcgagcgg	180
DB	102553	GCGGAGACGGTCGTCGGAACACGGCTCCCGGCATGGTAAACCGCGGGCGGAGCGG	102494
QY	181	gtcactttgcccatactcctacagaaactcccagaactccgcttccctagtccaacccaa	240
DB	102493	GTCTACTTTGCGGCAATCTCTACGAGAACTCCCAAGAACTCGCTTCCTAGTCCCAACCAA	102434
QY	241	gccagagttgccacacctaagatgagcgggggcg-----	277
DB	102433	GCCAGAGTTGCCCCACACCTAAATAATGGCGGGCGGAGTGGCGCGCGCCCTCTGGG	102374
QY	278	-----atgacacggcccccgcgggt	297
DB	102373	CGGGACCGGGGACTAGACGTGGCGGGGGCGGTGTCTATCGCCCCCGCCCGCCGGT	102314
QY	298	ccagccagctcgcccggggttcgggtgttcgggcccgcgtccctctctcgcaggt	357

Search completed: July 15, 2002, 13:56:39
Job time: 15364 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:24:01 ; Search time 56.81 Seconds
(without alignments)
967.814 Million cell updates/sec

Title: US-09-776-865-4
Perfect score: 2617
Sequence: 1 MKSPVSLAPSDGEGSDRT.....LFAKGEVQNWAISHQGHNRN 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2617	100.0	495	21	AA1980
2	2617	100.0	495	22	AA1981
3	2329	89.0	495	21	AA1982
4	2329	89.0	495	22	AA1983
5	2329	89.0	536	21	AA1984
6	2329	89.0	536	22	AA1985
7	2329	88.7	495	21	AA1986
8	1702	65.0	495	21	AA1987
9	1544	59.0	314	22	AA1988
10	1232	47.1	277	22	AA1989
11	1217	46.5	309	22	AA1990

12	991.5	37.9	559	22	ABB58701	Drosophila melanog
13	985.5	37.7	502	22	ABB60525	Drosophila melanog
14	985.5	37.7	502	22	ABB65873	Drosophila melanog
15	956.5	36.5	582	22	AAW79273	Human protein SEQ
16	946.5	36.2	516	22	ABB67013	Drosophila melanog
17	941	36.0	529	22	ABB63684	Drosophila melanog
18	932	35.6	567	22	AAO13870	Human polypeptide
19	931	35.6	560	17	AAW05148	Human brain sodium
20	931	35.6	560	19	AAW70500	Human sodium-lithi
21	895	34.2	194	22	AAW25685	Human protein sequ
22	859.5	32.8	481	22	ABB64204	Drosophila melanog
23	845	32.3	479	22	ABB62841	Drosophila melanog
24	835.5	31.9	496	22	ABB61407	Drosophila melanog
25	819	31.3	576	20	AAW88523	Eat-4 protein amin
26	798	30.5	465	22	ABB64710	Drosophila melanog
27	787	30.1	512	22	ABB60925	Drosophila melanog
28	784.5	30.0	497	22	ABB70142	Drosophila melanog
29	767.5	29.3	462	22	ABB70143	Drosophila melanog
30	760	29.0	493	22	ABB59580	Drosophila melanog
31	744	28.4	512	21	AAW29915	Arabidopsis thalia
32	744	28.4	519	21	AAW29914	Arabidopsis thalia
33	737	28.2	475	22	ABB67155	Drosophila melanog
34	737	28.2	491	22	ABB70144	Drosophila melanog
35	729	27.9	395	21	AAW29916	Arabidopsis thalia
36	715.5	27.3	560	22	ABB59401	Drosophila melanog
37	697	26.6	449	22	ABB58659	Drosophila melanog
38	696.5	26.6	524	22	ABB64935	Drosophila melanog
39	672	25.7	436	19	AAW78919	Human haemochromat
40	651	24.9	343	22	ABB12999	Novel human diagno
41	618	23.6	495	22	ABB64689	Drosophila melanog
42	611	23.3	124	22	AAW81307	Human AFP protein
43	608	23.2	256	22	ABG02032	Novel human diagno
44	603.5	23.1	401	19	AAW78920	Human haemochromat
45	603.5	23.1	401	19	AAW69971	Human sodium-depen

ALIGNMENTS

RESULT	1
AA1980	AA1980
ID	AA1980 standard; Protein; 495 AA.
AC	AA1980
DT	31-MAY-2000 (first entry)
DE	Sheep GBS toxin receptor (SP55).
KW	Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55; pathological vascularisation; cancer metastases; angiogenesis; sheep; neovascularisation; reperfusion injury; scarring; keloid; chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; endothelial cell proliferation; antibacterial; anticancer; anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
OS	Ovis sp.
Key	Location/Qualifiers
Domain	226..252
Domain	/note= "Outer boundary of transmembrane domain"
Domain	232..248
Domain	/note= "Inner boundary of transmembrane domain"
Domain	365..389
Domain	/note= "Outer boundary of transmembrane domain"
Domain	369..385
Domain	/note= "Inner boundary of transmembrane domain"
Domain	456..479
Domain	/note= "Outer boundary of transmembrane domain"
Domain	458..474
Domain	/note= "Inner boundary of transmembrane domain"
Domain	135..157
Domain	/note= "Outer boundary of transmembrane domain"

FT Domain 137..153
 FT /note= "Inner boundary of transmembrane domain"
 FT Domain 42..58
 FT /note= "Predicted transmembrane domain"
 FT Domain 328..345
 FT /note= "Outer boundary of transmembrane domain"
 FT Domain 328..344
 FT /note= "Inner boundary of transmembrane domain"
 FT Domain 390..407
 FT /note= "Outer boundary of transmembrane domain"
 FT Domain 390..406
 FT /note= "Inner boundary of transmembrane domain"
 FT Modified-site 97..100
 FT /note= "Putative amidation site"
 FT Modified-site 59..62
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 71..74
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 77..80
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 95..98
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 225..228
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 302..305
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 357..360
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 11..14
 FT /label= CK2_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 73..76
 FT /label= CK2_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 79..82
 FT /label= CK2_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 259..262
 FT /label= CK2_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 452..455
 FT /label= CK2_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 126..131
 FT /note= "Predicted myristylation site"
 FT Modified-site 142..147
 FT /note= "Predicted myristylation site"
 FT Modified-site 162..167
 FT /note= "Predicted myristylation site"
 FT Modified-site 172..177
 FT /note= "Predicted myristylation site"
 FT Modified-site 205..210
 FT /note= "Predicted myristylation site"
 FT Modified-site 209..214
 FT /note= "Predicted myristylation site"
 FT Modified-site 337..342
 FT /note= "Predicted myristylation site"
 FT Modified-site 386..391
 FT /note= "Predicted myristylation site"
 FT Modified-site 403..408
 FT /note= "Predicted myristylation site"
 FT Modified-site 423..428
 FT /note= "Predicted myristylation site"
 FT Modified-site 427..432
 FT /note= "Predicted myristylation site"
 FT Modified-site 17..19
 FT /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 37..39
 FT /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT Modified-site 55..57
 FT /label= PKC_phospho_site

FT Modified-site /note= "Predicted phosphorylation site"
 FT 73..75
 FT /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT 97..99
 FT Modified-site /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT 254..256
 FT /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT 269..271
 FT /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT 276..278
 FT /label= PKC_phospho_site
 FT /note= "Predicted phosphorylation site"
 FT 8..35
 FT /label= p56a
 FT /note= "Immunogenic peptide"
 FT 7..22
 FT /label= p55a
 FT /note= "Immunogenic peptide"
 FT 71..84
 FT /label= p57a
 FT /note= "Immunogenic peptide"
 XX WO200005375-A1.
 XX 03-FEB-2000.
 XX 22-JUL-1999; 99WO-US16676.
 XX 22-JUL-1998; 98US-0093843.
 XX (UYVA-) UNIV VANDERBILT.
 XX Hellerqvist CG, Fu C;
 XX WPI: 2000-205377/18.
 XX N-PSDB; AA250876.
 PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
 PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
 XX Claim 10; Page 22; 109pp; English.
 CC The present sequence is sheep GBS (group B beta-haemolytic
 CC streptococci) toxin receptor (SP55). Sheep GBS toxin receptor is an
 CC integral protein with seven transmembrane domains. Expression vectors
 CC comprising the coding region can be transformed into host cells to
 CC express GBS toxin receptor and its fragments. Detecting the receptor in
 CC tissues is used to diagnose pathological vascularisation, e.g. for
 CC detecting cancer metastases. GBS toxin receptors are useful for treating
 CC conditions associated with pathological angiogenesis or
 CC neovascularisation (specifically cancer, reperfusion injury, scarring
 CC during wound healing, keloids, chronic inflammation (rheumatoid
 CC arthritis or psoriasis) or neural injury), and to raise specific
 CC antibodies used for treating early onset disease. Inhibitors of this
 CC receptor are useful for treating pathological or hypoxia-induced
 CC endothelial cell proliferation and migration.
 XX Sequence 495 AA;

Query Match 100.0%; Score 2617; DB 21; Length 495;
 Best Local Similarity 100.0%; Pred. No. 5e-276;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKSPVSDIAPSDGEGSDRTPLQRAPRAPAPVCCSARYNIAFLSFFGFFVLSLRVNL 60
 DB 1 mkspsvdsldapsdgesdrtplqraprapapvccsaryniaflsffgffvlsrlvnl 60
 QY 61 SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTKKRWDAETQGWILGSFFGYI 120

Db	61	svalvdmvdsntaktndrtsecahsapikvlhnqtgkkyrwdactggwllgsffgyi	120
QY	121	ITQIPGGYVASRSGKLLGFGIFATAIETLTPLAADFGVALRALLEGEGVTYP	180
Db	121	itqipggyvasrsgklllgfifataiftlftplaadfgvalralelgvegvtyp	180
QY	181	AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNWTYYVFFGIVGIIW	240
Db	181	amhamsswappplerskllsisyagaqlgtvvsplsgvicymmwnwtvyffgivgiww	240
QY	241	FILWICLVSDTETHKTIPTPYKEYILLSLKNQLSSQKSPWPIMLKSLPLWAIIVVAHFS	300
Db	241	filwiclvsdtpethkttitpykeyillslknglssqskspwipmlksiplwaiivvahfs	300
QY	301	YNWTFYTLTLPTYMKEVLRNIOENGFLSAVPYLGWCMLSGQAADNLRARNWST	360
Db	301	ynwtfytltltpymkevrlrniengflsavpylgwcmilsgqaadnlrarwnfst	360
QY	361	LWVRRVFLSGLMIGPAIFLVAAGFIGCDYSLAVAFITISTTLGGFCSSGFSINHLDIAPS	420
Db	361	lwvrrvflslgmigpaiflvaagfigcdyslavafitisttlggfcsgfsinhldiaps	420
QY	421	YAGILLGINTFATIPGMIGPIIARSLTPENTIGEQTVFCIAAINVFGAIFTFLFAKG	480
Db	421	yagillgintfatipgmigpiiarsltpeentigewqtvfciaaainvfgaiftflfakg	480
QY	481	EVQNWAISSDHQGHNRN	495
Db	481	evqnwaissdhqghnrn	495
RESULT 2			
ID	AAE06519 standard; Protein; 495 AA.		
AC	AAE06519;		
DT	16-OCT-2001 (first entry)		
DE	Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) protein.		
KW	Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55; cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic;		
KW	prevention; attenuation; pathoangiogenic condition; cancer; scar;		
KW	wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;		
KW	keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;		
OS	Ovis sp.		
PH	Key		
FT	Region	Location/Qualifiers	
FT		8..22	
FT		/label= p55a_immunogenic_peptide	
FT		/note= "Fragment of extracellular domain of GBS toxin receptor"	
FT	Region	9..35	
FT		/label= p56a_immunogenic_peptide	
FT	Region	14..19	"Region of high hydrophilicity"
FT	Region	25..30	/note= "Region of high hydrophilicity"
FT	Region	75..80	/note= "Region of high hydrophilicity"
FT	Region	71..84	/label= p57a_immunogenic_peptide
FT		/note= "Fragment of intracellular domain of GBS toxin receptor"	
PN	W0200156598-A2.		
XX	09-AUG-2001.		
PD			
XX			

PF	02-FEB-2001; 2001WO-US03662.		
XX			
PR	02-FEB-2000; 2000US-0179870.		
XX			
PA	(UYVA-) UNIV VANDERBILT.		
XX			
PI	Hellerqvist CG;		
XX			
DR	WPI; 2001-488844/53.		
DR	N-PSDB; AAD10326.		
XX			
PT	Preventing or attenuating pathoangiogenic conditions e.g. cancer,		
PT	chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by		
PT	administering group B beta-hemolytic Streptococci toxin receptor or its		
PT	fragment -		
XX			
PS	Claim 8; Page 50-52; 52pp; English.		
XX			
CC	The present sequence is group B beta-haemolytic Streptococci (GBS)		
CC	toxin receptor protein, SP55 from sheep. The present invention relates		
CC	to a method for preventing or attenuating a patho-angiogenic condition		
CC	in a mammal which comprises administering to the mammal one or more GBS		
CC	toxin receptors or their immunogenic fragments to induce or maintain an		
CC	immune response to one of GBS toxin receptors. The method is useful for		
CC	preventing or ameliorating pathoangiogenic conditions such as cancer,		
CC	scarring during wound healing, gliosis during repair of nerve injury,		
CC	chronic wounds, keloids, reperfusion injury, rheumatoid arthritis,		
CC	atherosclerosis, osteoarthritis and psoriasis in mammals. The proteins		
CC	of the invention are also used as vaccines.		
XX			
SQ	Sequence 495 AA;		
Query Match 100.0%; Score 2617; DB 22; Length 495;			
Best local Similarity 100.0%; Pred. No. 5e-276;			
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MKSPVSDLAPSDGEGSDRTPLLRAPRAEPAPVCCSARYNLAFISFFGFFVLSLRVNL	60
Db	1	mkspsvdlapsdgegsdrtpllqrapraepapvccsarynlafisffgffvlsyrnl	60
QY	61	SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTGKKYRWDATQGWILGSFFYGYI	120
Db	61	svalvdmvdsntaktndrtsecahsapikvlhnqtgkkyrwdactggwllgsffgyi	120
QY	121	ITQIPGGYVASRSGKLLGFGIFATAIETLTPLAADFGVALRALLEGEGVTYP	180
Db	121	itqipggyvasrsgklllgfifataiftlftplaadfgvalralelgvegvtyp	180
QY	181	AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNWTYYVFFGIVGIIW	240
Db	181	amhamsswappplerskllsisyagaqlgtvvsplsgvicymmwnwtvyffgivgiww	240
QY	241	FILWICLVSDTETHKTIPTPYKEYILLSLKNQLSSQKSPWPIMLKSLPLWAIIVVAHFS	300
Db	241	filwiclvsdtpethkttitpykeyillslknglssqskspwipmlksiplwaiivvahfs	300
QY	301	YNWTFYTLTLPTYMKEVLRNIOENGFLSAVPYLGWCMLSGQAADNLRARNWST	360
Db	301	ynwtfytltltpymkevrlrniengflsavpylgwcmilsgqaadnlrarwnfst	360
QY	361	LWVRRVFLSGLMIGPAIFLVAAGFIGCDYSLAVAFITISTTLGGFCSSGFSINHLDIAPS	420
Db	361	lwvrrvflslgmigpaiflvaagfigcdyslavafitisttlggfcsgfsinhldiaps	420
QY	421	YAGILLGINTFATIPGMIGPIIARSLTPENTIGEQTVFCIAAINVFGAIFTFLFAKG	480
Db	421	yagillgintfatipgmigpiiarsltpeentigewqtvfciaaainvfgaiftflfakg	480
QY	481	EVQNWAISSDHQGHNRN	495
Db	481	evqnwaissdhqghnrn	495

RESULT 3	
AA45087	
ID	AA45087 standard; Protein; 495 AA.
XX	
AC	AA45087;
XX	
DT	31-MAY-2000 (first entry)
XX	
DE	Partial human GBS toxin receptor (HP55).
XX	
XX	Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KW	pathological vascularisation; cancer metastases; angiogenesis;
KW	neovascularisation; reperfusion injury; scarring; keloid;
KW	chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW	endothelial cell proliferation; antibacterial; anticancer;
KW	anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Region
FT	Location/Qualifiers
FT	14..20
FT	/note= "Region of high hydrophilicity used to design
FT	antigenic peptide"
FT	76..80
FT	/note= "Region of high hydrophilicity used to design
FT	antigenic peptide"
FT	26..30
FT	/note= "Region of high hydrophilicity used to design
FT	antigenic peptide"
XX	
PN	WO200005375-A1.
XX	
PD	03-FEB-2000.
XX	
PF	22-JUL-1999; 99WO-US16676.
XX	
PR	22-JUL-1998; 98US-0093843.
XX	
PA	(UYVA-) UNIV VANDERBILT.
XX	
PI	Hellerqvist CG, Fu C;
XX	
DR	WPI; 2000-205377/18.
XX	
DR	N-PSDB; AA250875.
XX	
PT	New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT	useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX	
PS	Claim 9; Page 22; 109pp; English.
XX	
CC	The present sequence is partial human GBS (group B beta-haemolytic
CC	streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
CC	protein with seven transmembrane domains. Expression vectors comprising
CC	the coding region can be transformed into host cells to express GBS
CC	toxin receptor and its fragments. Detecting the receptor in tissues is
CC	used to diagnose pathological vascularisation, e.g. for detecting cancer
CC	metastases. GBS toxin receptors are useful for treating conditions
CC	associated with pathological angiogenesis or neovascularisation
CC	(specifically cancer, reperfusion injury, scarring during wound healing,
CC	keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC	neural injury), and to raise specific antibodies used for treating early
CC	onset disease. Inhibitors of this receptor are useful for treating
CC	pathological or hypoxia-induced endothelial cell proliferation and
CC	migration.
XX	
SQ	Sequence 495 AA;
XX	
Query Match 89.0%; Score 2329; DB 21; Length 495;	
Best Local Similarity 86.3%; Pred. No. 1.2e-244;	
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;	

QY	1	MKSPVSDLAPSDGEGSDRTPLLQRAPRAEPAPVCCSARYNLAFSLFFGFVLYSLRVNL	60
DB	1	ms:spvrdlarndgeestdrtllpqapraeapvccsarynlailaffgfviyalrvnl	60
QY	61	SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNOTGKKYRDWDAETQGWILGSFFYGI	120
DB	61	svalvdmvdsntlednrtskacpehsapikvhhnqtgkkyqwdetqgwilgsffgyi	120
QY	121	ITQIPGGYVASRGGKLLGFGIFATAIFTPLAADFGVALVALRALEGLGEGVTYP	180
DB	121	itqipggyvasriggkmlilgfgilgtavltlftpiaadlgvplivlralelgevtfp	180
QY	181	AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNWTYVYFFGIVGIW	240
DB	181	amhamsswapplerskllsisyagaqlgtvslplsgliCYMMNWTYVYFFGtgiw	240
QY	241	FILWICLVSDTETHKTIPTYKEYILSLKQLSSQKSVPMUKSLPLWAIIVVAHFS	300
DB	241	filwlvdsdtpqkhkrihsykeyilslrnqlssqksvpmvplksiplwvvhfs	300
QY	301	YNWTFVTLTLPTYMKEVLRFNIQENGLSAVPYLGCWLCMLTSGAADNLRARNFST	360
DB	301	ynwtfvltltpctymkellrfnvqngflsslpylgswicmlsgaadnlrakwnfst	360
QY	361	LMWRRVFLSMIGTGAIFLVAAGFIGCDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS	420
DB	361	lcvrrifsligmipavflvaagfigcdyslavafiltisttlggfcsgfsinhldiaps	420
QY	421	YAGILLGINTFATPGMIGPIIARSLTPENTIGEQTVFCIAAINVGAIFTLFAKG	480
DB	421	yagillgitntfatpgmvgpviaksltpdntvgewqtvfyiaaainvgaifftlfakg	480
QY	481	EVQNNWAISSDHOGRN	495
DB	481	evqnnwaindhghrh	495
RESULT 4			
AA66967			
ID	AA66967 standard; protein; 495 AA.		
XX			
AC	AA66967;		
XX			
DT	18-APR-2001 (first entry)		
XX			
DE	Human AST.		
XX			
KW	Human; AST; nootropic; immunotropic; gene therapy; Salla disease;		
KW	anion and sugar transporter; anion-cation symporter;		
KW	sialic acid transporter.		
XX			
OS	Homo sapiens.		
XX			
PN	EP1069184-A1.		
XX			
PD	17-JAN-2001.		
XX			
PF	16-JUL-1999; 99EP-0202341.		
XX			
PR	16-JUL-1999; 99EP-0202341.		
XX			
PA	(ALKU) AKZO NOBEL NV.		
XX			
DR	WPI; 2001-193090/20.		
DR	N-PSDB; AAF55900.		
XX			
PT	New human transporter gene implicated in Salla disease and lysosomal		
PT	sialic acid transport, useful in assays for identifying new drugs, or		
PT	diagnosing sialic acid transport defects related to mutations in the		
XX	transporter gene		
XX			
PS	Claim 1; Page 10-12; 20pp; English.		
XX			

CC	The present sequence is human Anion and Sugar Transporter (AST) protein sequence. AST has significant homology with several members of the anion-cation symporter (ACS) family of transporters. AST is implicated in Salla disease, and is useful in screening assays for identifying new drugs. Compounds identified via AST screening is useful for preparing a pharmaceutical suitable as an activator or inhibitor of a sialic acid transporter protein. The pharmaceutical may be used in sialic acid associated diseases and CNS/immune related disorders.									
XX										
SQ	Sequence 495 AA;									
Query Match 89.0%; Score 2329; DB 22; Length 495;										
Best Local Similarity 86.3%; Pred. No. 1.2e-244;										
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;										
QY	1	MKSPVSLAPSDGEESDRTPLLQRAPRAPAPVCCSARYNLAFSGFEFVLYSLRVNL	60							
Db	1	mrspvrdlarndgeesdrplpgpraeeapvccsarynlallaifgflvyalrvnl	60							
QY	61	SVALVDVDSNTTAKDNRTSYECAHSAPIKVLHNOTGKRYMDAETQGWILGSFFYGYI	120							
Db	61	svalvdmvdsnttledrtscakcpehsapikvhhnqtgkyqdaetqgwilgsffgyi	120							
QY	121	ITQIPGGYVARSCKLLGFGIFAPAFITFLTPPLAADFCVGALVALRLEGIGEVYTP	180							
Db	121	itqipggyvaskiggmllgflgtavltlftpiaadlgvgplivlrallengigevgtfp	180							
QY	181	AMHAMSSWAPPLERSKLLSISVAGAQGVVSLPLSGVTCYVMNTYVFYFFGIVGIW	240							
Db	181	amhamsswapplerskllsisvagaqgvcvisipisglicyymnwtvyfffgtgifw	240							
QY	241	FILWICLVSDTPETHRTITPYKEYIYLSSLKNQLSSQKSPWPIPMKLSPLWAIVVAHFS	300							
Db	241	flilwlvsdtpqkhrishyekeyilsslirnlssqsksvpwpilksplwaivvahfs	300							
QY	301	YNWTFYTLTLTPYKVELRFNIQENGLSVPYLGCLWILSGQADNLRARNFST	360							
Db	301	ynwtfytlitltpymkelirfnvqngfllssipylgswlcmilsggaadnkrakwnfst	360							
QY	361	LWRRVFSLIGMTGPAIFLVAAGFIGCDYSLAVAFITISTLGGFCSSGFSINHLDIAPS	420							
Db	361	lvrrifsligmgaavfivaagfigcdyslavafitistctlggfcsgfsinhlidiaps	420							
QY	421	YAGILLGITNTFATIPGMIGPITARSILTPENTIGEQTVFCIAAANVFGAIFFTLFAKG	480							
Db	421	yagillgitntfatipgmvgpvlaaksltpdntvgewtvfciaaainvfgaifftlfakg	480							
QY	481	EVONWAIISDHQHRN	495							
Db	481	evqpnwalndhbhgrh	495							
RESULT 5										
AA45089										
ID	AA45089 standard; Protein; 536 AA.									
XX										
AC	AA45089;									
XX										
DT	31-MAY-2000 (first entry)									
XX										
DE	Human GBS toxin receptor (HP59).									
XX										
KW	Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;									
KW	pathological vascularisation; cancer metastases; angiogenesis;									
KW	neovascularisation; reperfusion injury; scarring; keloid;									
KW	chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;									
KW	endothelial cell proliferation; antibacterial; anticancer;									
KW	anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.									
XX										
OS	Homo sapiens.									
XX										
FH	Key	Location/Qualifiers								

FT	Modified-site	23..26	/note= "Putative amidation site"
FT	Modified-site	138..141	/note= "Putative amidation site"
FT	Modified-site	100..103	/note= "Asn is N-glycosylated"
FT	Modified-site	112..115	/note= "Asn is N-glycosylated"
FT	Modified-site	118..121	/note= "Asn is N-glycosylated"
FT	Modified-site	136..139	/note= "Asn is N-glycosylated"
FT	Modified-site	266..269	/note= "Asn is N-glycosylated"
FT	Modified-site	343..346	/note= "Asn is N-glycosylated"
FT	Modified-site	398..401	/note= "Asn is N-glycosylated"
FT	Modified-site	297..300	/label= CAMP_phospho_site
FT	Modified-site	113..116	/note= "Putative phosphorylation site"
FT	Modified-site	114..117	/label= CK2_phospho_site
FT	Modified-site	300..303	/note= "Putative phosphorylation site"
FT	Modified-site	493..496	/label= CK2_phospho_site
FT	Modified-site	66..71	/note= "Putative phosphorylation site"
FT	Modified-site	167..172	/note= "Putative myristylation site"
FT	Modified-site	183..188	/note= "Putative myristylation site"
FT	Modified-site	213..218	/note= "Putative myristylation site"
FT	Modified-site	246..251	/note= "Putative myristylation site"
FT	Modified-site	250..255	/note= "Putative myristylation site"
FT	Modified-site	378..383	/note= "Putative myristylation site"
FT	Modified-site	427..432	/note= "Putative myristylation site"
FT	Modified-site	444..449	/note= "Putative myristylation site"
FT	Modified-site	464..469	/note= "Putative myristylation site"
FT	Modified-site	468..473	/note= "Putative myristylation site"
FT	Modified-site	23..25	/label= PKC_phospho_site
FT	Modified-site	58..60	/label= PKC_phospho_site
FT	Modified-site	78..80	/note= "Putative phosphorylation site"
FT	Modified-site	120..122	/label= PKC_phospho_site
FT	Modified-site	138..140	/note= "Putative phosphorylation site"
FT	Modified-site	310..312	/label= PKC_phospho_site
FT	Modified-site		/note= "Putative phosphorylation site"

Modified-site 317..320
/label= PKC_phospho_site
/note= "Putative phosphorylation site"
WO200005375-A1.
03-FEB-2000.
22-JUL-1999; 99WO-US16676.
22-JUL-1998; 98US-0093843.
(UYVA-) UNIV VANDERBILT.
Hellerqvist CG, Fu C;
WPI; 2000-205377/18.
N-PSDB; AA250879.
New polynucleotide encoding mammalian receptor for streptococcus toxin,
useful for diagnosis and treatment of, e.g. pneumonia in neonates -
Claim 10; Page 93-95; 109pp; English.
The present sequence is partial human GBS (group B beta-haemolytic
streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
protein with seven transmembrane domains. Expression vectors comprising
the coding region can be transformed into host cells to express GBS
toxin receptor and its fragments. Detecting the receptor in tissues is
used to diagnose pathological vascularisation, e.g. for detecting cancer
metastases. GBS toxin receptors are useful for treating conditions
associated with pathological angiogenesis or neovascularisation
(specifically cancer, reperfusion injury, scarring during wound healing,
keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
neural injury), and to raise specific antibodies used for treating early
onset disease. Inhibitors of this receptor are useful for treating
cell pathological or hypoxia-induced endothelial cell proliferation and
migration.
Sequence 536 AA;
Query Match 89.0%; Score 2329; DB 21; Length 536;
Best Local Similarity 86.3%; Pred. No. 1.4e-244;
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKSPVSLAPSDGEGSDRPTLQRAPAPVCCSARYNLAPLSFFGFVLYSLRVNL 60
Db mrsprdlarndgeestdrtpllpgpraeapvccsarynlailatfgfviyalrvnl 101
QY 61 SVALVDMVDNNTAKNRTSYECAEHSAPIKVLHNOTGKKYRWDATQGWLTGSEFFGYI 120
Db 102 svalvdmvdnttdrtkscapsapikvhhmqgkryqwdacqgwlsgffgyi 161
QY 121 ITQIPGYVASRGGKLLGLGIFATPFTPLAADFGVAGLVALRALEGEGVYTP 180
Db 162 itqipgyvaskggkmlfgilgtavltlftpiaadlcvglvrlralelegvgtfp 221
QY 181 AMHAMSSWAPPLERSKLLSISAGAGLGVTVSLPLSGVICYNNWTVYFFPGVGIIV 240
Db 222 amhamsswappplerskllsisagaglgvtvslplsgliicyymnwvtyffgtgfw 281
QY 241 FILWICLVSDPTFKHTITPYEKEYILSLKLNLSQKSPWIPMLKSLPLWAIWAHFS 300
Db 282 flwivlsdtpqkhrishyekeyilslrnqlsqkspwvplkslplwaiwvahfs 341
QY 301 YNNFTYFLPTMYKEVLFNFQENGFLSVPYPLGWCMLISGQAADNLARNWFST 360
Db 342 ynnftytltpymkeilfnfvqengflsslpylgswclmllsgqaadnlrakwnfst 401
QY 361 LWRRVPSLLGMPAIFLVAAGFICGDISLAVAFILTISTLGGFCSSGSINHLDIAPS 420
Db 402 lcwrrfslgmgpavflvaagfigcdyslavafitistlggfcsggsinhldiaps 461

QY 421 YAGILGITTTFATIPGMIGPIIARSILTPENTIGEMOTVFCIAAAINVFAGIITLFAKG 480
Db 462 yagilgitttfatipgmvgpviaksltpdntvgewqtvfyaaainvfaiiftllfakg 521
QY 481 EVQNWAISSDHQHRN 495
Db 522 evqnwaindhghrh 536
RESULT 6
AAE06518
ID AAE06518 standard; Protein; 536 AA.
AC AAE06518;
XX
DT 16-OCT-2001 (first entry)
XX Human group B beta-haemolytic Streptococci toxin receptor (HP59) protein.
DE Human; group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;
XX cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Region 8..28 /label= Hab3_immunogenic_peptide
FT Region 49..63 /label= Hab1_immunogenic_peptide
FT Region 49..76 /label= Hab4_immunogenic_peptide
FT Region 112..125 /label= Hab2_immunogenic_peptide
XX
PN WO200156598-A2.
XX
XX 09-AUG-2001.
XX 02-FEB-2001; 2001WO-US03662.
XX 02-FEB-2000; 2000US-0179870.
XX (UYVA-) UNIV VANDERBILT.
XX Hellerqvist CG;
XX WPI; 2001-488844/53.
XX N-PSDB; AAD10325.
XX Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment -
XX Claim 4; Page 45-47; 52pp; English.
XX The present sequence is full length group B beta-haemolytic Streptococci
CC (GBS) toxin receptor protein, HP59 from human. The present invention
CC relates to a method for preventing or attenuating a patho-angiogenic
CC condition in a mammal which comprises administering to the mammal one
CC or more GBS toxin receptors or their immunogenic fragments to induce
CC or maintain an immune response to one of GBS toxin receptors. The
CC method is useful for preventing or ameliorating pathoangiogenic
CC conditions such as cancer, scarring during wound healing, gliosis
CC during repair of nerve injury, chronic wounds, keloids, reperfusion
CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
CC psoriasis in mammals. The proteins of the invention are also used
CC as vaccines.

XX	Sequence	536 AA:	
XX	Query Match	89.0%:	Score 2329; DB 22; Length 536;
XX	Best Local Similarity	86.3%:	Pred. No. 1.4e-244;
XX	Matches 427; Conservative	30;	Mismatches 29; Indels 0; Gaps 0;
QY	1	MKSPVSDLAPSDGEGSDRTPLLRAPRAEPAPCCSARYNLAFSLFFGFVLYSLRVNL	60
DB	42	mrspvrdlarndgeesdtrtlpgapraeapccsarynlalaffgfivayalrvnl	101
QY	61	SVALVDVDSNTTAKDNRTSYECAEHSAPIKVLHNQTKKRYDAETOGWILGSFFYGYI	120
DB	102	svalvdvdsnttlednrtskacpehsapikvhnhnqtgkygwdaetgwgilgsffgyi	161
QY	121	ITQIPGGVVASKGGKLLGLGFIATFTLTPLAADFGVGAIVLRALSGLGGVHYIP	180
DB	162	itqipggvvaskigkmlglfigltavltlftpiaadlgvppilviralsglgegvlfpp	221
QY	181	AMHAMWSSWAPPLERSKLLSYAQAQGLTVVSLPLSGVICYIMNMTVYVFFFGIVGIW	240
DB	222	amhamwsswapplerskllsisyagaqlgtvisplsgliwymnwtvvyfffgtgifw	281
QY	241	FILWICLVSDTPETHKTIPTPEKEYIYLSLKNQISSQKSPWIPMLKSLPLWAIYVAHFS	300
DB	282	filwivdsdtpqkhrishyekeyiylsslnqissqkspwvpilksplwailvvaahfs	341
QY	301	YNWTFYTLTLTPYMKVEVLRFNIOENGFLSAVPYLGOWCLMILSGQAADNLRARWNEST	360
DB	342	ynwtfytltltpymkeilrfnvngengfslslypgslwclmilsqgaadnkrakwnst	401
QY	361	LWRRVFSLIGMIGPAIFLVAAGFIGDYSIYLAFAFLTITTLGGFCSSGFSINHIDIAPS	420
DB	402	lcwriifslgmigpavflvaagfigdcyslavafitistlglgfcssgfsinhidiaps	461
QY	421	YAGILLGITNTFATIPGMIPGIARSITPENTIGEMQVFCIAAINVFGAIFPTLPAGK	480
DB	462	yagillgitntfatipgmvgpviaksitpdtntvgewqtvfyiaaainvfgalftlfakg	521
QY	481	EVQNWAITSDHOGHRN 495	
DB	522	evqnwaindhghrnh 536	
RESULT	7		
ID	AAM38959		
XX	AAM38959	standard; Protein; 495 AA.	
AC	AAM38959;		
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE	Human polypeptide	SEQ ID NO 2104.	
XX			
KW	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	anyotrophic lateral sclerosis; Shy-brager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX	Homo sapiens.		
OS			
XX			
XX			
PN	W0200153312-Al.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		

```

|||||
421 yagillgitntfatipgmvpviakstlpdntvgewqtvfyiaaainvgaiftlifakg 480
|||||
481 EVONWALSDHQHNRN 495
|||||
481 evqnwalndhghrh 495

RESULT 8
AA45090
ID AAY45090 standard; Protein; 495 AA.
AC AAY45090;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human/Sheep consensus GBS toxin receptor.
XX
KW Human/Sheep consensus GBS toxin receptor; GBS;
KW group B beta-haemolytic streptococci; pathological vascularisation;
KW cancer metastases; angiogenesis; neovascularisation; reperfusion injury;
KW scarring; keloid; chronic inflammation; rheumatoid arthritis; psoriasis;
KW neural injury; endothelial cell proliferation; antibacterial;
KW anticancer; anti-angiogenic; anti-inflammatory; anti-arthritis;
KW anti-psoriatic.
XX
OS Homo sapiens.
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT Protein 1..495
FT /note= "All the X's in the sequence correspond to
FT non-consensus sites"
XX
PN WO200005375-A1.
XX
XX 03-FEB-2000.
XX
XX 22-JUL-1999; 99WO-US16676.
XX
XX 22-JUL-1998; 98US-0093843.
XX
XX (UYVA-) UNIV VANDERBILT.
XX
XX Hellerqvist CG, Fu C;
XX
XX WPI; 2000-205377/18.
XX
XX N-PSDB; AA250880, AA250881.
XX
XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
XX Disclosure; Page 98-100; 109pp; English.
XX
XX The present sequence is the human/sheep consensus GBS (group B beta
XX -haemolytic streptococci) toxin receptor. Expression vectors comprising
XX the coding region can be transformed into host cells to express GBS
XX toxin receptor and its fragments. Detecting the receptor in tissues is
XX used to diagnose pathological vascularisation, e.g. for detecting cancer
XX metastases. GBS toxin receptors are useful for treating conditions
XX associated with pathological angiogenesis or neovascularisation
XX (specifically cancer, reperfusion injury, scarring during wound healing,
XX keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
XX neural injury), and to raise specific antibodies used for treating early
XX onset disease. Inhibitors of this receptor are useful for treating
XX pathological or hypoxia-induced endothelial cell proliferation and
XX migration.
XX
XX Sequence 495 AA;

```

Query Match 65.0%; Score 1702; DB 21; Length 495;
 Best Local Similarity 67.8%; Pred. No. 2.7e-176;

```

Matches 335; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 1 MKSPVSDLAPSGEGSDRTPLLQRAPEAPAPVCCSARYNLAFLSFFGFFVLYSRVNL 60
DB 1 mxpxvdxaxxxgeexdxrxxxxxaxrxexapccsarynaxlxfiffxxylxvnl 60
QY 61 SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNTGKKRWDAETQCWILGSPFYGYI 120
DB 61 vxxvxxmxdsxttxdnrxsxxcexhsapikvxxxqtkgkxwdaetqgwilxxfygyi 120
QY 121 ITQIPGGYVASRGGKLLGLGFIGIFATAIFTLFTPLAADFGVALVALRALEGIGEVYTP 180
DB 121 itxipggyvasxxgxxxlqxlxxxaxxtlftpxaadxxgxxxxlxalexlgegtxpx 180
QY 181 AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNWTYFYPFGIIGIIV 240
DB 181 amhamsswapplersxlxxiyagaxlgtvxsplsgvicymnwtvyfxxfxxxgxxw 240
QY 241 FILWICLVSDTPTHTKTITPYEKEYTLSSLKQLSOKSVPPWIPMLKSLPLAIVVAHFS 300
DB 241 fxxwixlvxtpxxhxxxyekxxllslxqxxsqkxvpxxxxxxlpwaxxvaxfs 300
QY 301 YNWFYTLTLPTMYKVELRFNIQENGLSAVPYLGWCMLSGQAADNLRARWNFT 360
DB 301 ynwftfytlxllptxmkxxlrfnxqngflsxxpylxxwlcmlxgqaadnlarxwnfst 360
QY 361 LWVRVFLSGLIGMIGPAIFLVAAGFIGCDYSLAVAFITITTLGGFCSSGFSINHLDIAP 420
DB 361 xxvxrxfslixmipxxflxxxxxxgdyxlxxvflxistxlggfcsgfsinhxiaps 420
QY 421 YAGILGINTTATIPGMIGPIIARSLTPENTIGEMQTVFCIAAANVFGAIFTFLFAG 480
DB 421 yagilgicnxfatixgmxxpxixsxtbntxgwxqxxfxaaaxnxfagixtlfakg 480
QY 481 EVONWALSDHQHNR 494
DB 481 exqpnwxxxdhxghr 494

RESULT 9
AAG65238
ID AAG65238 standard; Protein; 314 AA.
XX
AC AAG65238;
XX
DT 20-NOV-2001 (first entry)
XX
DE Human sodium dependent phosphate co-transfer protein 35.
XX
KW Human; sodium dependent phosphate co-transfer protein 35;
KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy.
XX
OS Homo sapiens.
XX
XX CN1298882-A.
XX
XX 13-JUN-2001.
XX
XX 06-DEC-1999; 99CN-0124217.
XX
XX 06-DEC-1999; 99CN-0124217.
XX
XX (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2001-503367/56.
XX
XX N-PSDB; AAH79234.
XX
XX Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX Claim 1; Page 21(Disclosure); 28pp; Chinese.
PS

```

XX The present invention provides the protein and coding sequences of human
 CC sodium dependent phosphate co-transfer protein 35. The sequences can be
 CC used in the treatment of hypophosphaturia, hypercalcaemia,
 CC hypophosphataemic rickets and nephritis. The present sequence is the
 CC protein of the invention.
 XX
 SQ Sequence 314 AA;

Query Match 59.0%; Score 1544; DB 22; Length 314;
 Best Local Similarity 89.2%; Pred. No. 2.3e-159;
 Matches 280; Conservative 24; Mismatches 10; Indels 0; Gaps 0;

Qy 182 MHAMWSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNTVYVFFGIVGIWF 241

Db 1 mhawsswappplerskllsisyagqlgtvislpslglicymnwtvvyffgtigifw 60

Qy 242 ILWICLVSDTETHKTIPTPYEKEYILSLKQLSSQKSPVWIPMLKSLPLWAIIVVAHFSY 301

Db 61 llwlvsvdtpqkhrshyekeyillrnlqslssqkspvwpilksiplwaiivvahfsy 120

Qy 302 NWTFTLLTLPTYMKEVLRNIQENGLSAVPYLGWLCMLTSGQAADNLRARWNFSTL 361

Db 121 nwtftlltlptymkeilrfnvqngflssipylgswlcmilsgqaadnhrakwnfstl 180

Qy 362 WYRRVSLIGMIGPAIFLVAAGFIGCDYSLAVAFLTITTLGGFCSSGFSINHLDIAPSY 421

Db 181 cvrrifsligmigpavflvaagfigcdyslavafitistlgtgfcssgfsinhldiapsy 240

Qy 422 AGILGINTFATPGMIGPIIARSLTPENIGEQVFCIAAINVFGAIFFTLFAKGE 481

Db 241 agilgintfatipgmvgpviaksltpdntvgewtvyfaaaainvfgaifftlfakge 300

Qy 482 VQWNAISDHOGRN 495

Db 301 vqnwaindhghnrh 314

RESULT 10

AAM93914
 ID AAM93914 standard; Protein; 277 AA.

XX AC AAM93914;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide, SEQ ID NO: 4069.

XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.

XX OS Homo sapiens.

XX PN EP1130094-A2.

XX PD 05-SEP-2001.

XX PF 07-JUL-2000; 2000EP-0114089.

XX PR 08-JUL-1999; 99JP-0194486.

XX PR 11-JAN-2000; 2000JP-0118774.

XX PR 02-MAY-2000; 2000JP-0183765.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX DR WPI; 2001-524255/58.

XX DR N-PSDB; AAK94876.

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -

XX

PS Claim 8; SEQ ID NO 4069; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesising full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

CC molecules have been determined. Primers for synthesising the full length

CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping

CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is a polypeptide

CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 277 AA;

Query Match 47.1%; Score 1232; DB 22; Length 277;

Best Local Similarity 81.9%; Pred. No. 1.9e-125;

Matches 227; Conservative 22; Mismatches 28; Indels 0; Gaps 0;

Qy 1 MKSPVSDIAPSDGEGSDRTPLQRPAPRAEPAPVCCSARYNLAFSLFFGFFVLYSLRVNL 60

Db 1 mrsprvdlarndgeestdrtpllpgapraeapvccsarynlalaffgfivayalrvnl 60

Qy 61 SVALVDMVDSNTTAKDNRTSYECAHSAPIKVLHNQTKYRWDATOGWILGSFFYGYI 120

Db 61 svalvdmvdsnttlednrtskacpshsapikvhhnqtgkkyqwdatogwilgsffyyi 120

Qy 121 ITQIPGGYVASKGKLLGFGIFATAFTLTPLAADFGVALVALRALEGLSGEVTP 180

Db 121 itqipggyvaskgkmlgfgilgtavltlftplaadlvgvpllvralleglsgvufp 180

Qy 181 AMHAMWSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNTVYVFFGIVGIW 240

Db 181 amhamwsswappplerskllsisyagqlgtvislpslglicymnwtvvyffgtigifw 240

Qy 241 FILWICLVSDTETHKTIPTPYEKEYILSLKQLSSQ 277

Db 241 filwlvsvdtpqkhrshyekeyillslrnqagvq 277

RESULT 11

AAM40745.

ID AAM40745 standard; Protein; 309 AA.

XX AC AAM40745;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 5676.

XX KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-052317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

QY 441 PIIARSITPENTIGEWQTVFCIAAAINVECAIFFFTLFAKGEVONWA 486
||| :|| : ||| || ||| :: | ||: :|| || | ||
Db 461 piivgamtknksreewqyvfliaalvhyggvifayafasgekpwa 506

Search completed: July 15, 2002, 08:24:02
Job time: 107 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:58 ; Search time 32.66 Seconds
(without alignments)
1456.345 Million cell updates/sec

Title: us-09-776-865-4
Perfect score: 2617
Sequence: 1 MKSPVSLAPSDGEGSDRT.....LFAKGEVQNWAISHQHRN 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	968.5	37.0	493	2 G88553	protein C38C10.2 [
2	950	36.3	472	2 S28286	hypothetical prote
3	927	35.4	560	2 I59302	brain specific Na+
4	819	31.3	563	2 T43650	probable sodium-de
5	819	31.3	576	2 H88548	protein ZK512.6 [i
6	809.5	30.9	465	2 A56410	sodium/phosphate t
7	802.5	30.7	465	2 S69915	sodium/phosphate t
8	792.5	30.3	573	2 T23589	hypothetical prote
9	791.5	30.2	467	2 A48916	sodium phosphate t
10	775.5	29.6	465	2 I39473	Na+-dependent phos
11	766.5	29.3	544	2 T24633	hypothetical prote
12	744	28.4	512	2 H84698	hypothetical prote
13	585	22.4	501	2 B89135	protein F25G6.7 [i
14	561.5	21.5	413	2 T01534	hypothetical prote
15	550.5	21.0	537	2 T45634	hypothetical prote
16	541.5	20.7	466	2 S40767	hypothetical prote
17	499	19.1	561	2 D84800	hypothetical prote
18	490.5	18.7	592	2 T25419	hypothetical prote
19	489	18.7	568	2 S44742	C02C2.4 protein -
20	471	18.0	530	2 T29418	hypothetical prote
21	450	17.2	380	2 T23248	hypothetical prote
22	444.5	17.0	516	2 T24729	hypothetical prote
23	428	16.4	499	2 T15201	hypothetical prote
24	420	16.0	506	2 T23968	hypothetical prote
25	419.5	16.0	543	2 T32496	hypothetical prote
26	395.5	15.1	493	2 T25357	hypothetical prote
27	391.5	15.0	445	2 T23590	hypothetical prote
28	382.5	14.6	425	2 A90055	hypothetical prote
29	379.5	14.5	473	2 T31717	hypothetical prote

30	373	14.3	455	1 H69752	probable glucarat
31	371	14.2	485	2 T24115	hypothetical prote
32	369	14.1	452	2 AE0861	probable glucarate
33	368.5	14.1	478	2 T33942	hypothetical prote
34	366.5	14.0	462	2 T34365	hypothetical prote
35	365.5	14.0	420	2 S44900	ZK52.10 protein -
36	361	13.8	450	2 A65061	probable glucarate
37	358.5	13.7	428	2 AD0675	membrane transport
38	357	13.6	450	2 A91085	probable transport
39	357	13.6	450	2 C85930	probable transport
40	346.5	13.2	455	2 T34366	hypothetical prote
41	339.5	13.0	659	2 T33557	hypothetical prote
42	338.5	12.9	422	1 A69853	hexuronate transpo
43	334.5	12.8	493	2 T19383	hypothetical prote
44	331.5	12.7	516	2 T27092	hypothetical prote
45	326	12.5	487	2 T23776	hypothetical prote

ALIGNMENTS

RESULT 1

G88553

protein C38C10.2 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001

C:Accession: G88553

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C-elegans/ and www.sanger.ac.uk/projects/C-

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: G88553

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <STO>

A:Cross-references: GB:chr_III; PIDN:CAA79549.1; PID:g3874873; GSPDB:GN00021; CESP:C

C:Genetics:

A:Gene: C38C10.2

A:Map position: 3

Query Match 37.0%; Score 968.5; DB 2; Length 493;

Best Local Similarity 40.5%; Pred. No. 2.6e-70;

Matches 193; Conservative 89; Mismatches 179; Indels 15; Gaps 6;

QY 27 PRAEPAPVCCSARYNAFLSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRTSYECAH 86

DB 8 PRLVP-----STRFALSLVMFFGCLVYMMRTNMKSAFVCMVNMENKTDGTGVEKVSRCGE 62

QY 87 SAPIKVLHNOTGKKYRWDATOGMILGSFFYGIITQIPGGYVASRSGGKLLGFGIFAT 146

DB 63 MTPVESNSSVIG-EFWDQKOTTMGVLSFFYGIIGHLASRYGKRRVVFVTLGS 121

QY 147 AFTLTPLTAADFGVALVALRALEGIGVITYPAMHAMSSWAPPLERSKLLISYAGA 206

DB 122 ALLTLNPVAARTSEYALAILRAAIGFLOGATFPAHMTKMSVWGPPLLSVLTVGYAGA 181

QY 207 QLGTVVSLPLSGVICVY---MNVTVYFFGIVGIITWILICVLSDDTETHKTIPTPEK 263

DB 182 QIGNVIVLPLSGFLCEYFGDGGWPSIFYIIGVFGVLWTAVMWYVSSDKPATHPRITPEEK 241

QY 264 EYTLSSLKQWLSQ---KSVPHIPMLKSLPLWAIIVAHFSYNWTFYTLTLTPTVMKEV 319

DB 242 QYIVTVAASMGKDTGKVFPSTPWIKILTSPAVWACWAGHFGAGWGYATMLVSPSLKDV 301

QY 320 LRENIENGFLSAVYILGWLCLMILSCQAADNLRARWNFTLWVRVFSLIGMIGPAIFL 379

DB 302 LGNLSSLGAVASIPYIAFLAINAGGVLDATLRSKGLSTLNTRRAMVALIGOGIFL 361

QY 380 VAAGFTGCCDYS-LAVAFLTISTTTLGGFCSSGGSINHLDTAPSAGILLGTNTFATPGM 438

Db 362 VASGYCGGQDVLIIFITCGMAISLQYAGFVVVNYLEIAPPFGSTVMGTGNTISALAGI 421
QY 439 IGPIIARSLTPENTIGEWQTFVFCIAAAINVEGAIFFTFLFAKGEVQNWNA-ISDHQGH 493
Db 422 ISPAVSYLTPNGTQEBQWVNLWTAGILLIGALLFSIFASGEVQWPWAKUTAEEGH 477

RESULT 2
S28286
hypothetical protein C38C10.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 12-Mar-1993 #sequence_revision 02-Jul-1996 #text_change 16-Feb-1997
C:Accession: S28286
R:Thomas, K.
submitted to the EMBL Data Library, December 1992
A:Reference number: S28285
A:Accession: S28286
A:Molecule type: DNA
A:Residues: 1-472 <THO>
A:Cross-references: EMBL:Z19153
C:Genetics:
A:Introns: 50/3; 287/3; 351/3; 412/3
C:Keywords: transmembrane protein

Query Match 36.3%; Score 950; DB 2; Length 472;
Best Local Similarity 41.0%; Pred. No. 7.6e-69;
Matches 187; Conservative 86; Mismatches 173; Indels 10; Gaps 5;

QY 47 FFGFVLYSLRVNLSVALVMDVSNNTAKNRTSYECAHSAPIKVLHNTGKRYRDAE 106
Db 2 FFGCLVTYMMRTNNSFAVVCWVNNKTDGTGVEKYSRCGKEMTPVESNSSVIG-EFDWDKQ 60
QY 107 TQGWILGSFFGYXIIITQPGSYVASRSGKLLGLGFGIFATAIETLFTPLAADFGVGA 166
Db 61 TTGVLVSSFFYIGISGIIQGHLSARVGRVVFVITLGSALLTLNPNVAARTSEYALAI 120
QY 167 LRALEGGEVGYTPAMHAMSSWAPPLERSKLLSISVAGAQGLGVTVSLPSGVICYV--- 223
Db 121 LRAAIGLQCATGPAMHTMSVNGPPLLSVLTVGVYAGAQIGNVIVLPSGLCEYGF 180
QY 224 MNVTYVFFGIVGIWFIWILWICLVSDTPETHTKTITPYEKEYILSSLKNOSSQ----KS 279
Db 181 GMPSPFIYIIGVFGVLWTAVMYVSSDKPATPRITPEEKQYIVTVAEASMGKDTGKVP 240
QY 280 VPMPLKSLPLWAIVVAHFSYNTWTFVTLTLTPTYMKEVLRNIOENGFLSAVPVLCW 339
Db 241 TPWIKILTSPAVWACHAGHPAGWGAYTMLVSLPSFLKDVGLNLSLGAVALSIPIYAF 300
QY 340 LCMILSQQAADNLRARNFNSTLWRRVFSLIGMIGPAIFLVAAGFTGCDYS-LAVAF 398
Db 301 LAI NAGGLADTLRSKILSTLNTTRAAMLVALIGQIFLVASGYCGGDVLIIFITC 360
QY 399 STTILGGFCSSGFNSHLDIAPSVAGILLGTNTFAIPGMIGPIIARSLTPNTIGEWQT 458
Db 361 GMAISLQYAGFVVVNYLEIAPPFGSTVMGTGNTISALAGISPAVSYLTPNGTQEBQW 420
QY 459 VFCIAAAINVEGAIFFTFLFAKGEVQNWNA-ISDHQGH 493
Db 421 VLMLTAGILLIGALLFSIFASGEVQWPWAKUTAEEGH 456

RESULT 3
I59302
brain specific Na+-dependent inorganic phosphate cotransporter - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C:Accession: I59302
R:Ni, B.; Rosteck, P.R.; Nadi, N.S.; Paul, S.M.
Proc. Natl. Acad. Sci. U.S.A. 91, 5607-5611, 1994
A:Title: Cloning and expression of a cDNA encoding a brain-specific Na(+)-dependent ino
A:Reference number: I59302; MUID:94261635
A:Accession: I59302

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-560 <RES>
A:Cross-references: EMBL:U07609; NID:g507414; PIDN:AAA19646.1; PID:g507415

Query Match 35.4%; Score 927; DB 2; Length 560;
Best Local Similarity 39.4%; Pred. No. 6.6e-67;

Matches 184; Conservative 89; Mismatches 168; Indels 26; Gaps 4;

QY 30 EPAPVCCSA-----RYNLAFSLFFGFVLYSLRVNLSVALVMDVSNNTAKNRTSYECA 84
Db 48 DPPVWDCTCFLPRRYIIATIMSGLCISFGICRNGLGVAIVSMVNNSTHRRGHVVVQKA 107
QY 85 EHSAPIKVLHNTGKRYRDAETQGWILGSFFGYXIIITQPGSYVASRSGKLLGLGIF 144
Db 108 Q-----FNMDPETVGLIHGFFWGYIVTQIPGGFICOKFAANRVFGFAIV 152
QY 145 ATAIFTLTPLAADFGVALVRALEGLGEGVTVTPAMHAMSSWAPPLERSKLLSISA 204
Db 153 ATSTLNLIPSAARVHYGCVIFVRILQGLVEGVTYPACHGINSKWAPPLERSLATATFC 212
QY 205 GAOLGTVWSLPSGVICYNNNTYVFFYFVGIWFIWILWICLVSDTPETHTKTITPYEKE 264
Db 213 GSYAGAVVAMPLAGVLVQYSGSSVFYVYGSFGIFWYLFWLLVSYESPALHPSISEERK 272
QY 265 YILSSLKNOSSQSV-----PWIMPKSLPLWAIVVAHFSYNTWTFVTLTLTPYMKEV 319
Db 273 YIEDALGESAKLMNPYTKFNTPMWRREFTSMPYAIIVANFCRSWTFYLLLSIQPAYFEV 332
QY 320 LRNIQENGFSLSAVYPLGWCMLSCQAADNLRARNFNSTLWRRVFSLIGMIGPAIFL 379
Db 333 FGEIISKVGLVSALPHLVMTIIVPIGQIADFLSRHIMSTTNVRLMNCGGFGMEATLL 392
QY 380 VAAGFTGCDYSLAVAFLTISTTILGGFCSSGFNSHLDIAPSVAGILLGTNTFAIPGM 439
Db 393 LVVGY-SHSGKGAISFLVLA VGSFGAISGFNVNHLDIAPRYASILMGISNGVGTLSGW 451
QY 440 GPIIARSLTPENTIGEWQTFVFCIAAAINVEGAIFFTFLFAKGEVQNWNA 486
Db 452 CPIIVGAMTKHKTREEMQYVFLIASLVHYGGVIFYGVFASGEKQPWA 498

RESULT 4
T43650
probable sodium-dependent inorganic phosphate cotransporter - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T43650
R:Lee, R.Y.N.; Sawin, E.R.; Chalfie, M.; Horvitz, H.R.; Avery, L.
submitted to the EMBL Data Library, September 1998
A:Description: EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate c
A:Reference number: Z22599
A:Accession: T43650
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-563 <LEE>
A:Cross-references: EMBL:AF095787; PIDN:AAC64972.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: eat-4
A:Map position: 3
C:Function:
A:Description: is necessary for glutamatergic neurotransmission

Query Match 31.3%; Score 819; DB 2; Length 563;
Best Local Similarity 36.3%; Pred. No. 3.4e-58;
Matches 165; Conservative 88; Mismatches 178; Indels 24; Gaps 6;

QY 36 CSARYNLAFSLFFGFVLYSLRVNLSVALVMDVSNNTAKNRTSYECAHSAPIKVLIN 95
Db 53 CRKRWLLAILANNGFMISFGIRCNGCAKTHMYKNYTD-----PYGKVHM 97

Qy	96	QTGKKYRWDAETGOWILGSGFFGYIITQIPGGVYASRSKGKLLGLGFATAIFTLTPL	155
Db	98	H--EFNWTIDELSVMESSFYGYLVTQIPAGFLAAKFPNKLFGFGVGAFLNILLPY	154
Qy	156	AADFGVCAVA-LRALEGLEGVITYPAMHAMSSWAPPLERSKLLSISVAGAQLGTVYSL	214
Db	155	GFKVSDYLAFTQITQGLVQGVCPAMHGVMWRYWAPPERSKLATATGTSYAGAVGL	214
Qy	215	PLSGVICYNNWTVYFFFGIVGIIWFIWILMICLVSDTPETHKITIYPEREYI--LSSLK	271
Db	215	PLSAFLSVYSWAAPFYLGVCVGIWAILWFCVTFEKPAPHFTISOEKFIEDAIGHVS	274
Qy	272	NQLSSOKSPVPIIDMLKSLPLWALVVAHSYNTFTYLLPLPTYMKEVURFNIQENGFLS	331
Db	275	NTHPTIRSIPIKAIATSKPWAIIIVANFARSFTYLLQNLQNTYKMAELGMIADSLIA	334
Qy	332	AVPYLCWCLMILSGQAADNLRAWNFSILWVRVPSLIGMTGPAIFLVAAGFIGCDYSL	391
Db	335	AIPHVLVWGCVLMGGQLADYLRNKLSTAVAKRIFNCGGFGGEAFMLVIAYTTSD-TT	393
Qy	392	AVAFALTSTTLGGFCSSGFSINHLDIAPSVAGILLGITNTFTATIPQMIGPIIARSUTPEN	451
Db	394	AIMALAAVCKSGFATSGFNVNHLDIAPRYAILMGFSNGIGTLAGLTCPFTVEAFTAHS	453
Qy	452	TIGEWOTVFCIAAIVFGAIFETLFAKEGVQNWAA	486
Db	454	KHG-WTSVFLASLIHFTGVTTFVAVYASGELQEWAA	487
RESULT	5		
H88548		protein ZK512.6 [imported] - Caenorhabditis elegans	
C:Species:		Caenorhabditis elegans	
C:Date:		10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2000	
C:Accession:		H88548	
R:anonymous,		The C. elegans Sequencing Consortium.	
Science		282, 2012-2018, 1998	
A:Title:		Genome sequence of the nematode C. elegans: a platform for invest	
A:Reference		number: A75000; MUID:9069613; PMID:9851916	
A:Note:		see websites genome.wustli.edu/gsc/C_elegans/ and www.sanger.ac.uk/	
A:Note:		published errata appeared in Science 283, 35, 1999; Science 283, 2	
A:Accession:		H88548	
A:Status:		preliminary	
A:Molecule		type: DNA	
A:Residues:		1-576 <STO>	
A:CROSS-references:		GB:chr_III; PIDN:CRAA80150.1; PID:g3881690; GSPDB:GN0000	
C:Genetics:			
A:Gene:		ZK512.6	
A:Map position:		3	

Db 288 NTHPTKSIIPKWAIVTSKPVWALIVANFARSWTFYLLDQLNQLTYMKEALGKMTADSGLLA 347
 Qy 332 AVPYLGCWLCMILSGQAADNLRARWNPSTLWVRVFLSLIGMIGPAIFLVAAGFIGDYSL 391
 Db 348 AIPHYMGCVVLMGQLADYLRNKKILSTTAVRKIFNCGGEGEAAFMILVAVYTTSD-TT 406
 Qy 392 AVAFLLTSTTLGGFCSSGFSINHLDTAPSYAGILLGTTWTFTATIPGMGIIARSLTPEN 451
 Db 407 AIMALIAAAGMSGFAISGFNVNHLDTAPRYAAILMGFSNGIGTLAGLTCPFVTEAFTAH 466
 Qy 452 TIGEWQTVFCIAAAINVFGAIFFTLFAKGEVQNA 486
 Db 467 KHG-WTSVFLASLIHFTGVTFAVYASGELQEWA 500

 RESULT 6
 A56410
 sodium/phosphate transport protein, renal - rabbit
 N:Alternate names: sodium/phosphate cotransporter
 C:Species: *Oryctolagus cuniculus* (domestic rabbit)
 C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
 C:Accession: A56410; S27951
 R:Werner, A.; Moore, M.L.; Mantel, N.; Biber, J.; Semenza, G.; Murer, H.
 Proc. Natl. Acad. Sci. U.S.A. 88, 9608-9612, 1991
 A:Title: Cloning and expression of cDNA for a Na/P-i cotransport system of kidney cor
 A:Reference number: A56410; MUID:92052140
 A:Accession: A56410
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-465 <WER>
 A:Cross-references: GB:M76466; NID:gl65689; PIDN:AAA31461.1; PID:gl65690
 C:Keywords: kidney; transmembrane protein

 Query Match 30.9%; Score 809.5; DB 2; Length 465;
 Best Local Similarity 36.0%; Pred. No. 1.6e-57;
 Matches 164; Conservative 86; Mismatches 196; Indels 9; Gaps 2;

 Qy 33 PVCCSARYNIAFISGFFGVLYSLRVNLVALVDVDSNTAKDNRTSYSCAEHSAPIKV 92
 Db 11 PCFSFRVYLALEMHFCNIVIIAQMCLMTWAMV-----NNTNLHGSPNTSAER 62
 Qy 93 LHNQTGKKYRWDAAETCGWILGSPFYGIITQPGVYVASRGGKLLGFGIFATAIPTLF 152
 Db 63 LDNTKPNVYNWSPDVQGLIFSSIFYGAFLLQIPGVYISGYSIKKLLGFAFLUSSLSIF 122
 Qy 153 TPLAADFGVALRALALEGLEGVTVTPAMHAMSSWAPPLERSKLLSISYAGAQLGTVV 212
 Db 123 IPQAAANGETWIIICRVVQGITQGTVTTAQHEIWWKAPPLERCLTSMLSGFLGPF 182
 Qy 213 SLPLSGVICYKMYNWTYVYFVGIVGIITWILWICLVSDTPTHTKITPYKKEYILSSLKN 272
 Db 183 VLLVTGIIICESLGPWMPYIFYGACGACVLLIVLYYDDPKDHPCVSLHEKEYITSSLIQ 242
 Qy 273 QLSS-QKSVPIWPKLSPLWAIWVAHFSYNWTFYLLTLPTMKEVLRFNQENGFLS 331
 Db 243 QGSSTROSLPTKAMIKSLPLWAISSFCFAYLWTVSRIVTPTLLINSMHLVDIRENGLS 302
 Qy 332 AVPYLGCWLCMILSGQAADNLRARWNPSTLWVRVFLSLIGMIGPAIFLVAAGFIGCDYSL 391
 Db 303 SLPYLFAWICGVIAGHTADFLMSRNLSLTAIRKLFTAIGLLLPVFSMCLLYLSSGFYS 362
 Qy 392 AVAFLLTSTTLGGFCSSGFSINHLDTAPSYAGILLGTTWTFTATIPGMGIIARSLTPEN 451
 Db 363 TITELIANANSSSPCLGALINALDAPRYVYFTKGVTTLTGMTGGTTSSTVAGLFLSQD 422
 Qy 452 TIGEWQTVFCIAAAINVFGAIFFTLFAKGEVQNA 486
 Db 423 PESSWFKIFLMSIINVISVIFYLIFAKAEIQDWA 457

S69915

sodium-phosphate transport system 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S69915
R:Chong, S.S.; Kozak, C.A.; Liu, L.; Kristjansson, K.; Dunn, S.T.; Bourdeau, J.E.; Hughes, A.M. J. Physiol. 268, 1038-1045, 1995
A:Title: Cloning, genetic mapping, and expression analysis of a mouse renal sodium-dependent anion channel protein
A:Reference number: S69915
A:Accession: S69915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-465 <CHO>
A:Cross-references: EMBL:X77241; NID:g887522; PIDN:CAA54459.1; PID:g887523

Query Match 30.7%; Score 802.5; DB 2; Length 465;
Best Local Similarity 36.5%; Pred. No. 5.8e-57;
Matches 166; Conservative 89; Mismatches 191; Indels 9; Gaps 2;
QY 33 PVCCSARYNLAFSLSEFGFVLYSLRVNLVALVDVDSNTTAKDNRTSYECAHSAPIKV 92
DB 11 PGCSFRYGLAILLHFCNTAINAQRCLNLTVMVMV-----NNTGSPHLSNESVEM 62
QY 93 LHNQTKKYRWDAETQGWILGSGFFYGIITQIPGVVASRSGKLLGLGFGIFATAITLFL 152
DB 63 LDNVKNPVYSWSPDIOGLILSSVFFGCMVVVQAPVGLSGIYPMKRIIGSLFLSSLSLL 122
QY 153 TPLAADFGVALVALRALGEGVITYPAMHAMSSWAPPLERSKLLSYAGAGLGTVV 212
DB 123 IPPAAQVGAALVTCRVLOIAQGVSTGQHEIWKWAPPLERGRUITSMTLSGFVMGPI 182
QY 213 SLPLSGVICYNNWTVYVFFGVLGIWILWICLVSDTPETHKTIPTPYEKEYILSLKN 272
DB 183 VLVSGFICDLGWPVYIFGIVGCVLSWFFLEFDDPKDHPYSSSEKDYIIISLMQ 242
QY 273 QLSS-QKSPWIPMLKSLPLWAIWVAHFYNWTFYLLTLLPTMYKEVLRNFNIQENGFLS 331
DB 243 QAASSGROSLPIKAMKSLPLWAILLSFAFINSLSLVYTPFTFISTVLHVNVRENGLS 302
QY 332 ANPILGCVLCMILSGQAADNLARNWFLWVRVSLGMIGTPAIFLVAAGFIGDYSL 391
DB 303 SLPYLLAYICGLLAGOMSDFFLTKRIFSIVTKRKLFTLGSFCPVIFIMCLLYLSNFYS 362
QY 392 AVAFLRISTTLGGSCSGFSINHLDIAPSAGVILLGITNTFATIPGMIGPIIARSLTPEN 451
DB 363 TVIFLILANSTLSFSYCGQLINALDIPRYGFLKAVTALIGMFGGLISSTLAGLILNOD 422
QY 452 TIGEWTVFCIAAANVFGAIFFTLFAKGEVQNW 486
DB 423 PEVAHKISFLMAGINVTCLVFFFLFAKGEIQDWA 457

RESULT 8
T23589
hypothetical protein K10G9.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23589; T24636
R:Wortimore, B.
submitted to the EMBL Data Library, August 1994
A:Reference number: 219765
A:Accession: T23589
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <WIL>
A:Cross-references: EMBL:Z36282; PIDN:CAA85289.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone K10G9
R:Buck, D.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19915
A:Accession: T24636
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-573 <WIL>A:Cross-references: EMBL:Z48055; PIDN:CAA88135.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone T07A5

C:Genetics:

A:Gene: CESP:K10G9.1

A:Map position: 3

A:Introns: 38/3; 87/3; 224/1; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match 30.3%; Score 792.5; DB 2; Length 573;
Best Local Similarity 37.1%; Pred. No. 4.7e-56;
Matches 171; Conservative 85; Mismatches 164; Indels 41; Gaps 9;

QY 39 RYNLAFLSFGFVLYSLRVNLVALVDVDSNTTAKDNRTSYECAHSAPIKVLINQGT 98
DB 47 RWOIALLAHFGFAISFGIRSNFGVAKNRMINNFTDAYE-----VHE--- 88
QY 99 KKYRWDAETQGWILGSGFFYGIITQIPGVVASRSGKLLGLGFGIFATAITLFTPLAAD 158
DB 89 KEFFWTGTEVGMMESSFFYGYAASQIPAGVIAAKFAPNKLFLMLGILFASLLNIVTAICLN 148
QY 159 FGVA---LVALRALGEGVITYPAMHAMSSWAPPLERSKLLSYAGAGLGTVVSLP 215
DB 149 FHPFTDIFVMVIVQMQLGALVCYPAMHGVMKYWAPPLERSKLATTTFTGASVGMVGLP 208
QY 216 LSGVICYNNWTVYVFFGVLGIWILWICLVSDTPETHKTIPTPYEKEYILS-----SL 270
DB 209 ASAYLVSHFSWSTPFYVFGALGIVMSILWIYVSGTSPETHGYISADEKKYITEKVSVA 268
QY 271 KNLSSOKSVPTPMLKSLPLWAIWVAHFYNWTFYLLTLLPTMYKEVLRNFNIQENGFL 330
DB 269 KNM--TLTILPWRDMTSTAVWAIICSCRSWSFLLGNQLTYMKDVLHDIKNSGLI 326
QY 331 SAVPYLGCVLCMILSGQAADNLARNWFLWVRVSLGMIGTPAIFLVAAGFIGC--- 387
DB 327 AIFPQLGMCIVTLTSGQLSDYLRSKGMSTEAVRKSVNTFG-----FTVEAVMLGCLAF 380
QY 388 --DYSLAVAFITISTTLGGSCSGFSINHLDIAPSAGVILLGITNTFATIPGMIGPIIAR 445
DB 381 VRDPVIAVTLIITACSGAGAVLSGFNVNHFDAIRHAPILMGIANGLAGIAG-VGGIVTN 439
QY 446 SLTPNTIGEWTVFCIAAANVFGAIFFTLFAKGEVQNW 486
DB 440 SLTYQNPDG-WQWVFLANSLIDIFGIIFLFAKGDVLPWA 479

RESULT 9

A48916

sodium phosphate transport protein 1, renal - human

C:Species: Homo sapiens (man)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999

C:Accession: A48916

R:Chong, S.S.; Kristjansson, K.; Zoghbi, H.Y.; Hughes, M.R.

Genomics 18, 355-359, 1993

A:Title: Molecular cloning of the cDNA encoding a human renal sodium phosphate transp

A:Reference number: A48916; MUID:94117004

A:Accession: A48916

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-467 <CHO>

A:Cross-references: GB:X71355; NID:g450531; PIDN:CAA50490.1; PID:g450532

C:Genetics:

A:Gene: GDB:SLC17A1; NPT1

A:Cross-references: GDB:I41889; OMIM:182308

A:Map position: 6p23-6p21.3

Query Match

Best Local Similarity 30.2%; Score 791.5; DB 2; Length 467;

Matches 180; Conservative 79; Mismatches 183; Indels 39; Gaps 7;

QY 23 LQRAPRAEP--APVCCSARYNLAFISFFGFFVLYSLRVNLVALVDVDS-----NTT 73

Db 1 MDMNRLLPKKPGGCSFRIGLSELVHCCNVITTAQRACLNLTVMVWVNSTDPHGLPNTS 60
QY 74 AKDNRTSYECASHPAPIKVLHNOTGKRYWDAETOCWILGSGFFGYIITQIPGGYVASRS 133
Db 61 TK-----KLLDNKPMYNWSPDIOGIISSYSYGVIIIOVPVGFSGIY 105
QY 134 GKLLGFGIFATAITFTPLAADFGVALVRALEGLGEGVTYPAMHAMSSWAPPL 193
Db 106 STKKMIGFALCLSSLLIPPAAGTGAWVVVCRVQAAQIVATAQAEIYVVKWAPPL 165
QY 194 ERSKLLSYAGAOLGTVVSLPLSGVICYMMWTVYFFFGIVGIWFIWILICLVSDTPE 253
Db 166 EGRITSMSTGFLGPPFTLVLTGVICESLGMWVYIFGACGCAVCLLWFVLFYDDPK 225
QY 254 THKTTTPYEKILSKNOL--SSOKSVPIWMLKSLPLWAIWVAHFSYNWTFYLLTLL 312
Db 226 DHPICISEKEVITSLVQVSSRSQSLPIKAILKSLPWAISISGFTFFWSHNMTLYT 285
QY 313 PTYMKVLRNFNIQENGLFSAVPLGCLWMLSGOAGNLRLARWNFSTLWVRVFSLIGM 372
Db 286 PMFINSMLHVNKENGFLSSLPYFAWICGNLAGOLSDFLTRNLSLVIAVRKLFATAAGF 345
QY 373 ICPAFLVAAGFIGCDYSLAVAFILSTTLGFCSSGFSINHLDTAPSYAGILLGITNTF 432
Db 346 LLPAIFGVCLPYLSFTYSIVIFLAGATGSCFCLGGVFINGLDIAPRYFGFI----KAC 401
QY 433 ATIPGMIGPIARSLT-----PENTIGEWOTVFCIAAAINVFCAIFFTLFAKGEVQNW 485
Db 402 STLTGMIGGLIASTLTGLLKODPESA---WFKTFILMAAINVTGLIFLIVATAEIQDW 458
QY 486 A 486
Db 459 A 459
RESULT 10
I39473
Na+-dependent phosphate cotransporter - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C:Accession: I39473
R: Miyamoto, K.; Tatsumi, S.; Sonoda, T.; Yamamoto, H.; Minami, Y.; Takeda, Y.;
Biochem. J. 305, 81-85, 1995
A: Title: Cloning and functional expression of a Na(+)-dependent phosphate co-transporter
A: Reference number: I39473; MUID: 95126933
A: Accession: I39473
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-465 <RES>
A: Cross-references: GB:D28532; NID:g639841; PIDN:BAA05888.1; PID:g639842

Query Match 29.6%; Score 775.5; DB 2; Length 465;
Best Local Similarity 37.3%; Pred. No. 8.7e-55;
Matches 175; Conservative 79; Mismatches 178; Indels 37; Gaps 6;
QY 33 PVCCSARYNLAFSFFGFFVLSLRVNLVALVDMVDS-----NTTAKDNRTSYECAE 85
Db 11 PGFCFRGLSLVHCCNVITTAQMLCLNLTVMVWVNSTDPHGLPNTSK----- 60
QY 86 HSAPIKVLHNOTGKRYWDAETOCWILGSGFFGYIITQIPGGYVASRSGLLGLFGIPA 145
Db 61 -----KLLDNKPMYNWSPDVQIILSSYSYGVIIIOVPVGFSGIYTKMKWIFALCL 115
QY 146 TAIFTLPTLAADFGVALVRALEGLGEGVTYPAMHAMSSWAPPLSKLLSISVAG 205
Db 116 SSVLSLIPPAAGTGAWVVVCRVQAAQIVATAQAEIYVVKWAPPLGRITSMSTSG 175
QY 206 AOLGTWVSLPLSGVICYMMWTVYFFFGIVGIWFIWILICLVSDTPEHKTITPYEKY 265
Db 176 FLLGPPFVLLVTGVICESLGMWVYIFGACCAVCLLWFLVFDPPKDHPCISIGEKY 235

QY 266 ILSSLKNOQL--SSOKSVPIWMLKSLPLWAIWVAHFSYNWTFYLLTLLPTYMKVLRNFI 324
Db 236 ITSSLVQVSSRSQSLPIKAILKSLPWAISISGFTFFWSHNMTLYTDMFINSMLHVN 295
QY 325 OENGFLSAVPLGCLWMLSGOAGNLRLARWNFSTLWVRVFSLIGMIGPAIFLVAAGF 384
Db 296 KENGFLUSSLPYLFAMTCGNLAGOLSDFLTRNLSLVIAVRKLFATAAGFLLPAIFGVCLPY 355
QY 385 IGCYDSLAVAFILSTTLGFCSSGFSINHLDTAPSYAGILLGITNTFTATPQMIGPIIA 444
Db 356 LSTFYSIVIFLILAGATGSCFCLGGVFINGLDIAPRYFGFI----KACSTLTGMIGGLIA 411
QY 445 RSLT-----PENTIGEWOTVFCIAAAINVFCAIFFTLFAKGEVQNW 486
Db 412 STLTGLILKODPESA---WFKTFILMAAINVTGLIFLIVATAEIQDW 457
RESULT 11
T24633
hypotheetical protein T07A5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24633
R: Buck, D.
submitted to the EMBL Data Library, February 1995
A: Reference number: Z19915
A: Accession: T24633
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-544 <WIL>
A: Cross-references: EMBL:Z48055; PIDN:CAA88134.1; GSPDB:GN00021; CESP:T07A5.3
A: Experimental source: clone T07A5
C: Genetics:
A: Gene: CESP:T07A5.3
A: Map position: 3
A: Introns: 38/3; 87/3; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match 29.3%; Score 766.5; DB 2; Length 544;
Best Local Similarity 35.6%; Pred. No. 5.5e-54;
Matches 164; Conservative 90; Mismatches 166; Indels 41; Gaps 9;
QY 39 RYNLAFPLSPFGFFVLSLRVNLVALVDMVDSNTTAKDNRTSYECASHPKVLHNOTG 98
Db 47 RQCIATLAHFGPAISGIRSNFVAKRMVNNFTDAYGE-----VHE--- 88
QY 99 KKYRWDAAETOGWILGSGFFGYIITQIPGGYVASRSGLLGLFGIFATAIFTLFTPLAAD 158
Db 89 REFLLWTGAEVGMWESSFFGYAASQIPAGVLAAKFAPNKFIMLGILVASFMNLSAISFN 148
QY 159 FGVGA---LVALRALEGLGEGVTYPAMHAMSSWAPPLSKLLSISYAGAOLGTVVSLP 215
Db 149 FHPYTDIFVWVQAVOGLALGVLYPAMHGVKFWAPPLSKLATAFTFGSSVGVMTGLP 208
QY 216 LSGVICYMMWTVYFFFGIVGIWFIWILICLVSDTPEHKTITPYEKYILS-----SL 270
Db 209 ASAYILVSHESWSPPFYVFGVGIWLSLIWVYSSHPETHGVISDDEKKQVTEKIGDVAV 268
QY 271 KQOLSSOKSVPIWMLKSLPLWAIWVAHFSYNWTFYLLTLLPTYMKVLRNFNIQENGL 330
Db 269 KNN--SLTTLPHRDMMTSSAVWAIITCTFCRSGWFFLLGNLTQMKVDLHIDIRKSGFI 326
QY 331 SAVPYLGCWMLSGOAGNLRLARWNFSTLWVRVFSLIGMIGPAIFLVAAGFICG--- 387
Db 327 SIFPQGMCIIVLATQGLCDYLRSSCKMSTEAVRKSNTFTG-----FIVEAMMLCCLAF 380
QY 388 --DYSIAVAFILSTTLGFCSSGFSINHLDTAPSYAGILLGITNTFTATPQMIGPIIA 445
Db 381 VRDPVIAVTCVLIACVSGVSLGSGFNHFDIAPRYATILMGLIANGLAGAVAG-VGGMVTIN 439
QY 446 STPENTIGEWOTVFCIAAAINVFCAIFFTLFAKGEVQNW 486
Db 440 TVTYQNPQG--WKWVFLAMAIIDIFGVIFLIFAKGDVLPWA 479

RESULT 12

H84698
hypothetical protein At2g29650 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84698
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: H84698
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:AE002093; NID:g3582333; PIDN:AAC35230.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29650
A:Map position: 2

Query Match 28.4%; Score 744; DB 2; Length 512;
Best Local Similarity 33.9%; Pred. No. 3.3e-52;
Matches 166; Conservative 83; Mismatches 187; Indels 54; Gaps 9;

QY 2 KSPVSDLAPSDGEGSDRTPLLQAPRA-----EPAPVCCSARYNLAFLSPFGFVLVS 55
DB 63 KSESSSYVEGDKVSGNNDVSDSPSSVLPWMEFP----KRWIVILCFSAFLCNM 118
QY 56 LRYNLVALVDVDSNTAKDNRTSYECAHSAPIKVLHNQTKYKRWDAETOGWILGSF 115
DB 119 DRYNMSIALP-----SAEYGNPATVGLIQSSF 148
QY 116 FYGIITQIPGGVVASRSGKLLGRGIFATAFTFTPLAADFGVGVALVALRALEGLGE 175
DB 149 FWGYLLTQIAGGIWADTVGGRVGLFGVIWSTATILTPVAAKLGFLPYLLVVRAEVGVGE 208
QY 176 GVTPYPAHAWMSWAPPLERSKLLSIYAGAQIGTVVSLPSGVICYVYNNWTVVYFFGI 235
DB 209 GVAMPAMNNILSKWVPQERSLALVSGMYLGSVTVGLAFSPFLIHQFGWPSVYFSGS 268
QY 236 VGIWFLWILCLVSDRTPTHTKTYPEKEYILLSSKNQLSOKSVPMIPMLKSLPLWAI 295
DB 269 LGIVWLTWLTAKESPLEDPILLPERKLIADNCASK-EPVKSIPWRLLSKPPVWALI 327
QY 296 VAHFSYNNWTFYLLTLPTMYKEVLRNIOENGFLSAVPLYGCWLCMLSGQA-----ADN 351
DB 328 SCHFCNMGTFILLTWMTYYHOVLKFNLMESGLLSVFP----WMTMAISANAGGWIADT 383
QY 352 LRARNFSLWVRVSLTGMIGPAIFLVAAGFIGDCYSLAVAFLLISTTLGGFCSSGFS 411
DB 384 LVSR-GFSTVNRKIMQITGFLGPAPFLQLKHIDSP-TMVLVCMACSGQTDFAFSQGLY 441
QY 412 INHLDIAPSVAGLLGITNTFATPGMIGPIIARSLTPENTIGEMOTVFCIAAINVEGA 471
DB 442 SNHQDIAPRYSGVLLSLNTAGVLGTAATGCHILQH---GSWDDVTISVGLILVGT 498
QY 472 IFTFLPAKGE 481
DB 499 VIWNLFSTGE 508

RESULT 13

B89135
protein F25G6.7 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: B89135
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: B89135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-501 <STO>
A:Cross-references: GB:chr_V; PIDN:AAC25800.1; PID:g2384837; GSPDB:GN00023; CESP:F25G
C:Genetics:
A:Gene: F25G6.7
A:Map position: 5

Query Match 22.4%; Score 585; DB 2; Length 501;
Best Local Similarity 29.9%; Pred. No. 2.1e-39;
Matches 149; Conservative 101; Mismatches 201; Indels 48; Gaps 13;

QY 4 PVSIDLAPSDGEGSDRTPLLQOR-APRAEPAPVCCSARYNLAFLSFFGFVLYSLRVNLSV 62
DB 6 PISDGIP-----SQQRPLTYRWF-----SWREMTSIMLCFCFCGCVHLMNSMGM 50
QY 63 ALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTKGKRWDAETOGWILGSFFYGIIT 122
DB 51 AIYCMVNSSAT-YDNET---YPENAAPL-----LDWSSDEQGYIFSAFNAGLLVM 96
QY 123 QIPGGVVASRSGK--LLLGFGIFATAFTFTPLAADFGVGVALVALRALEGLGEGVTP 180
DB 97 LFTGG-MADKFNKAKYMLVSLVAISLANFTL--PWAPISVYWAIFSRFLVGFADALLQP 153
QY 181 AHAMWSSWAPPLERSKLLSIYAGAQIGTVVSLPSGVIC----YYMNTVYVYFFGI 236
DB 154 AMNSLITRWFPTSESRVALGATGGQIGTLLIIPAGALCSQTEIFGGWPSFIFLSGFI 213
QY 237 GIWFLWILCLVSDRTPTHTKTYPEKEYILLSSKNQLS-----SOKSVPMIPMLKSLP 290
DB 214 GVLFIFSYIFLGADKPSKQSCISDNELKF--TISQSEVDGKKRTERKVPKMNILKSGA 271
QY 291 LWAIVVAHFSYNNWTFYLLTLPTMYKEVLRNIOENGFLSAVPLYGCWLCMLSGQAAD 350
DB 272 VVASVLSLVCHPEPLMTLWFLPSYLDVHHYHSTENGILSALPTVSLFAKIGSSYLNT 331
QY 351 NLRARNFSLWVRVSLTGMIGPAIFLVAAGFIGDCYS-LAVAFLLISTTLGGFCSSG 409
DB 332 WLQKNTTWKDKTICKVLNSIGSIGLGVFLAATFLDNEHAMVAVFLCLSMASAGLHTPG 391
QY 410 FSNHLDIAPSVAGLLGITNTFATPGMIGPIIARSLTPENTIGEMOTVFCIAAINVF 469
DB 392 QALALSVAPAYSGAVTGTFEFVAVSGIHPITIKMIVKDRTAAEWNLVFIYSTVIAIF 451
QY 470 GAIFFTLFAKGEVONWAS 488
DB 452 PIIFNVWGSTEVDWAKS 470

RESULT 14

T01534
hypothetical protein A_IG005110.nn - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 17-Mar-2000
C:Accession: T01534
R:Andrews, S.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IG005110.
A:Reference number: Z14347
A:Accession: T01534
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <CAND>
A:Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252847
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4

A; Introns: 79/3; 115/3; 128/3; 200/3; 288/3; 331/3; 353/3; 382/1
A; Note: A_IG005110.nh
C; Superfamily: hexuronate transporter

Query Match 21.5%; Score 561.5; DB 2; Length 413;
Best Local Similarity 28.1%; Pred. No. 1.3e-37;
Matches 145; Conservative 69; Mismatches 139; Indels 163; Gaps 12;

QY 2 KSPVSDL---APS-DGE-----EGS--DRTPLLRAPRAEPAPVCCSARYNLAFLSFF 48
DB 21 KSESDITECVPSADGSAEAIIVENLQNASPWQOQFPR-----RWIVILCF 70
QY 49 GFVFLVSLRNLVALVMDVDSNTAKDNRTSYECAHSAPIKVLHNQTKYKRYDAEQ 108
DB 71 SFLCNMDRVNMSIALPM-----SQEYNWSSATV 100
QY 109 GWILGFFGYIITQIPGGVVASRSGKLLLGFGIFATAIFTFTPLAADFGVALVALR 168
DB 101 GLIOSFFGWLTLQILGGIADKFGK----- 128
QY 169 ALRGLGEGVTPAMHAMWSWAPPLERSKLLSISYAGAOGLTVVSLPLSGVICVYNNWTY 228
DB 129 -----GVAMPANNMLSKWIPVSRSLALVYSGMILGSTGLAFSPMLITKFCWPS 181
QY 229 VFYFGIVGIWFIWICLVSDTPETHKTIPTPYEKEYIILSSKNQLSQKSPWPIMPLKS 288
DB 182 VFYFSGSLGSIWFLWLKFAYSPPKDDPLSEBEKKVILGSKPR-EPVTVIPWKLILSK 240
QY 289 LPLWALVVAHFSYNNWTFYLLTLPLPY-----MKEVLRFNIQENGL 330
DB 241 PPVWALIIFHCNHWGTFILLTWMPYYNQARSSASVISLLENFICEQVLKFNLTESGLL 300
QY 331 SAVPYLGLCWLMLGSOAANLRRNWFSLWVRRVFLSGLMIGPAIFLVAAGFIGCDYS 390
DB 301 CVLPWLTMAVFANIGWIADTLVSRG----- 326
QY 391 LAVAFITSTTLGGFCSSGFSINHLDIAPSYAGILLGITNTFATIPGMIGP-----IIAR 445
DB 327 -----LSITNGSDAFSQSGLYSNHODIGPRYAGVLLGSLNTAGVLGAGVGTAAATGYILQR 381
QY 446 SLTPENTIGEWQVFCIAAINVFGAIFFTLFAKE 481
DB 382 -----GSDDDVFKVAVALYLGTLVWNLFTATGE 409

RESULT 15
T45634
hypothetical protein F13112.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45634
R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, November 1999
A:Reference number: Z23010
A:Accession: T45634
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <CHO>
A:Cross-references: EMBL:AL133292
A:Experimental source: cultivar Columbia; BAC clone F13112
C:Genetics:
A:Map position: 3
A:Introns: 35/2; 153/3; 229/2; 350/3; 399/3; 445/3; 457/1
A:Note: F13112.30

Query Match 21.0%; Score 550.5; DB 2; Length 537;
Best Local Similarity 32.4%; Pred. No. 1.4e-36;
Matches 126; Conservative 75; Mismatches 155; Indels 33; Gaps 8;

QY 103 WDAETGWLGSFFGYIITQIPGGVVASRSGKLLLGFGIFATAIFTFTPLAADFGVG 162

Db 144 WSKSFGSIVQSSFLWGLYISPIAGGLTVDRYGGKVVWANGVALWSLATFTTPWAADSSLW 203
QY 163 ALVALRALEGLGEGVTPAMHAMWSWAPPLERSKLLSISYAGAOGLCTVVSPLSGVICY 222
Db 204 ALLAAAMVGVAGVALPCNNVAVRFPPTERSRAGVIAAGFQIGNVVGLMLSPILMS 263
QY 223 YMNWTVYVFFGVIGIWFILWICLVSDTPETHKTIPTPYEKEYI-----LSSKNQLSQ 277
Db 264 QGGIYGPVIFGLSGFLWLLWLSATSSAPDRHPQITKSELEYIKOKKQISTMENKRIST 323
QY 278 KSV-PWIPMLKSLPLWAIWVAHFSYNNWTFYLLTLPTYNKEVLRFNIQENGLSAVPYL 336
Db 324 SGIPPFGRLLSKRMTWAVIVANSMHSMVYHV-----NLKQAMFSAVP-- 366
QY 337 GCWLCMLISQAAA---DNLRRWNFSLWVRRVFLSGLMIGPAIFLVAAGFIGCDYSL-A 392
Db 367 --WSMAFTGYIAGFWSDDLIRGTSITLTKIMQSGIFGPGIALI--GLTTAKOPLVA 422
QY 393 VAFITSTTLGGFCSSGFSINHLDIAPSYAGILLGITNTFATIPGMIGPIIARSLTPENT 452
Db 423 SAWLSLAVGLKSFSLGLFLINLQEIAPESGVLHGMCLTAGTLAAIVGTGAGFFV--EL 480
QY 453 IGEWQIVFCIAAINVFGAIFFTLFAKE 481
Db 481 LGSFGQFILLTALYLLSALFYNIYATGE 509

Search completed: July 15, 2002, 08:22:59
Job time: 44 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:25:26 ; Search time 17.92 seconds
(without alignments)
1069.541 Million cell updates/sec

Title: US-09-776-865-4

Perfect score: 2617

Sequence: 1 MKSPVSDLAPSGEGSDRT.....LFAKGEVQNAISDHQHRN 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	968.5	37.0	493	YLD2_CABEL	Q03567 caenorhabdi
2	819	31.3	576	YQ06_CABEL	P34644 caenorhabdi
3	809.5	30.9	465	NPT1_RABIT	Q28722 o renal sod
4	802.5	30.7	465	NPT1_MOUSE	Q61983 mus musculu
5	797.5	30.5	465	NPT1_RAT	Q62795 rattus norv
6	790.5	30.2	465	NPT1_HUMAN	Q14916 homo sapien
7	766.5	29.3	544	YRT3_CABEL	Q10046 caenorhabdi
8	672	25.7	436	NPT3_CABEL	Q00624 homo sapien
9	603.5	23.1	401	NPT4_HUMAN	O00476 homo sapien
10	489	18.7	568	YKH4_CABEL	P34272 caenorhabdi
11	373	14.3	455	GUDP_BACSU	P42237 bacillus su
12	361	13.8	450	GUDP_ECOLI	Q46916 escherichia
13	338.5	12.9	422	EXUT_BACSU	Q34456 bacillus su
14	320.5	12.2	430	DGOT_ECOLI	P31457 escherichia
15	314	12.0	435	YFBO_BACSU	P37489 bacillus su
16	300	11.5	444	GARP_ECOLI	P42613 escherichia
17	291	11.1	453	XJL1_ECOLI	P39398 escherichia
18	281.5	10.8	472	EXUT_ECOLI	P42609 escherichia
19	261.5	10.0	456	GUDP_PSEPU	P42205 pseudomonas
20	246	9.4	429	YFAP_ECOLI	P76470 escherichia
21	226	8.6	433	TUB4_AGRVI	Q44470 agrobacteri
22	226	8.6	449	TUB3_AGRVI	P70786 agrobacteri
23	201.5	7.7	444	GLPT_BACSU	P37948 bacillus su
24	199	7.6	407	YB04_HAEIN	P71369 haemophilus
25	199	7.6	498	YAO1_SCHPO	Q10097 schizosacch
26	195.5	7.5	452	GLPT_ECOLI	P08194 escherichia
27	190	7.3	451	PHT1_PSEPU	Q05181 pseudomonas
28	189.5	7.2	480	GLPT_HAEIN	P96335 haemophilus
29	188	7.2	533	GLPT_HUMAN	P57057 homo sapien
30	187.5	7.2	413	MUCK_ACICA	P94131 acinetobact
31	184.5	7.1	403	MHPT_ECOLI	P77589 escherichia
32	174.5	6.7	427	YDFJ_ECOLI	P77228 escherichia
33	169	6.5	496	NANT_ECOLI	P41036 escherichia

RESULT 1

ID	YLD2_CABEL	STANDARD:	PRT:	493 AA.
AC	Q03567;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical 52.7 kDa protein C38C10.2 in chromosome III.			
GN	C38C10.2.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RX	MEDLINE=94150718; PubMed=7906398;			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coultson A.,			
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,			
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,			
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,			
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,			
RA	Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,			
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,			
RA	Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,			
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,			
RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,			
RT	Wohlman P.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans".			
RL	Nature 368:32-38(1994).			
RN	[2]			
RP	REVISTONS.			
RA	Jones S.J.M.;			
RA	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
RL	-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z19153; CAA79549.1; -			
DR	PIR; S28286; S28286.			
DR	WormPep; C38C10.2; CE08647.			
DR	InterPro; IPR003662; sub-transporter.			
DR	Pfam; PF00083; sugar tr; 1.			
KW	Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;			
KW	Sodium transport.			
FT	TRANSMEM 10 30			POTENTIAL.
FT	TRANSMEM 85 105			POTENTIAL.
FT	TRANSMEM 112 132			POTENTIAL.

34	166	6.3	412	1	YWFA_BACSU	P39637 bacillus su
35	163	6.2	457	1	PCAK_ACICA	Q43975 acinetobact
36	161.5	6.2	456	1	UHPT_CHLMU	Q9pj18 chlamydia m
37	160	6.1	572	1	YB30_YEAST	P38125 saccharomyc
38	159.5	6.1	456	1	UHPT_CHLTR	O84548 chlamydia t
39	158	6.0	405	1	YJHB_ECOLI	P39352 escherichia
40	158	6.0	451	1	YKJL_BACSU	P37514 bacillus su
41	155	5.9	448	1	PCAK_PSEPU	Q51955 pseudomonas
42	154.5	5.9	543	1	DAL5_YEAST	P15365 saccharomyc
43	151.5	5.8	388	1	NORA_STAAM	P21191 staphylococ
44	149.5	5.7	396	1	SOTB_ECO57	P58529 escherichia
45	149.5	5.7	396	1	SOTB_ECOLI	P31122 escherichia

ALIGNMENTS

```
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 205 225 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 441 461 POTENTIAL.
SQ SEQUENCE 493 AA; 52737 MW; 60792715D32553DD CRC64;

Query Match 37.0%; Score 968.5; DB 1; Length 493;
Best Local Similarity 40.5%; Pred. No. 2.9e-62;
Matches 193; Conservative 89; Mismatches 179; Indels 15; Gaps 6;

QY 27 PRAEPAPVCCSARYNLAFLSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRFTSYCAEH 86
DB 8 PRLPV-----STRFALSVMFFGCLVTYMMRTNMSPAVVCMVNEKTDGTGVEKVSRCGE 62

QY 87 SAPIKVLHNOTGKKYRWDATOGWILGSPFYGYIITQIPGGYVASRSGKLLGLGIFAT 146
DB 63 MTPVESNSSVIG-EFDWDKQTGMVLSFFYIGYSQIIIGHLSRYGGRVVFVTLGS 121

QY 147 AIFLTPLAADFGVALVALALEGLGEGVTPYPAHAMWSSWAPPLERSKLISISYAGA 206
DB 122 ALLTLNPAARTSEYALAILRAAIGFLOGATFPAMHTWMSVWGPPLLSVLGTGYAGA 181

QY 207 QLGTVVSLPLSGVICYI---MNTYVYFPGIIVTILICLVSDTPETHKTTIPYEK 263
DB 182 QIGNVTLPVLSGLFCEYFDGWPFSFYIIGVGLWTAVWVYSSDKPETHPRITPEEK 241

QY 264 EYTLSSLKWOLSSQ----KSPWIPMLKSLPLVAIVVAHFSYNWTPYTLTLPTVMKEV 319
DB 242 QYIVTAVEASMGKDTGKVPSTPIKILTPSPVWACHAGHFGAGWGAYTLMVLSPLSKDV 301

QY 320 LRFNIQENGLFSAVPYLGWLCMLSGQAADNLRARWNFTLWVRVFLSLIGMIGPAIFL 379
DB 302 LGLNLSLGAVASIPYIAVFLAINAGVGLADTLRSKGILSTLNTTRAAMVALIGQIFL 361

QY 380 VAAGFTGCDYS-LAVAFLTISTLGGFCSSGFSINHLDIAPSYAGILLGITNTFATIPQM 438
DB 362 VASGYCGCGQDVLVIFITCGMAISGLQYAGFVWVNYLEIAPPFSGTVMGTNTISALAGI 421

QY 439 IGPILARSITPENTIGEWTFVFCIAAINVFGAIFTLFAKEGVONWA-ISDHGH 493
DB 422 ISPAVSSYITPNTGQEWQWMLVLTAGILTGALLSFASGEVQPAKLTAEHG 477
```

```
RESULT 2
Y006 CAEEL STANDARD; PRT; 576 AA.
AC P34644;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 52.2 kDa protein ZK512.6 in chromosome III.
GN ZK512.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulcon L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
```

```
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22177; CAA80150.1; -.
DR PIR; S40767; S40767.
DR WormPep; ZK512.6; CE01109.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00883; sugar_tr; 1.
KW Hypothetical protein; Transmembrane; Glycoprotein; Symport;
KW Sodium transport.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 438 458 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
SQ SEQUENCE 576 AA; 63108 MW; 71F3A9EFBE5F84CC CRC64;

Query Match 31.3%; Score 819; DB 1; Length 576;
Best Local Similarity 36.3%; Pred. No. 1.7e-51;
Matches 165; Conservative 88; Mismatches 178; Indels 24; Gaps 6;

QY 36 CSARYNLAPLSPFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRFTSYECAHSAPIKVLHN 95
DB 66 CKRMLLAILANGFMISFGIRCNFCAKTHMYKNYTD-----PYGKVHM 110

QY 96 QTKKRYRWDATQGGWILGSPFYGYIITQIPGGYVASRSGKLLGLGIFATIFLTPL 155
DB 111 H---EFNWTIDELSVNMESSFYGYLVLTQIPAGFLAAKPPNKLFGGIGVGAFLNLLPY 167

QY 156 AADFGVGALVA-LRALEGLEGVTPYPAHAMWSSWAPPLERSKLISISYAGIQLGTVSVL 214
DB 168 GPKVRSQDLVAFITQIGLVQGVCPYPAHGVWRYWAPPMPERSKLATTAFTGSYAGAVLGL 227

QY 215 PLSGVICVYMMWTVYFPGIIVTILICLVSDTPETHKTTIPYEKYL---LSSLK 271
DB 228 PLASLVSTVSWAAPFYLYGCVGIWAILWFCVTEKFAFHFTISQEEKIFIEDHIGHS 287

QY 272 NOLSSQKSPVTPMLKSLPLVAIVVAHFSYNWTFYTLTLPTVMKEVLRFRNIQENGLS 331
DB 288 NTHPTIRSPKAIYTSKPVMAIIVANFARSWTFYLLQLNQLTYKKEALGMKIADSGLLA 347

QY 332 AVPYLGCWLCMLTSLGQAADNLRARWNFTLWVRVFLSLIGMIGPAIFLVAAGFICDYSL 391
DB 348 AIPHLVGVGVLMGGQLADYLRNKLSTLTAVRKIFNCGGFGGGAFAAFMLIVAYTTSDD-TT 406

QY 392 AVAFITISTLGGFCSSGFSINHLDIAPSYAGILLGITNTFATIPMIGPIIARSITPBN 451
DB 407 AIMALIAAVGMSGFAISGFNVNHLDIAPRYAALLGFGFNGICITLAGLTCPPFTTEAFTAH 466

QY 452 TIGEWQTVFCIAAAINVFGAIFTLFAKEGVONWA 486
```



```
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 337 356 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 429 449 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 465 AA; 51589 MW; C67E25A2C291EEF CRC64;

Query Match 30.7%; Score 802.5; DB 1; Length 465;
Best Local Similarity 36.5%; Pred. No. 2.1e-50;
Matches 166; Conservative 89; Mismatches 191; Indels 9; Gaps 2;

QY 33 PVCSARYNLAFLSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRTSYECABHSAPIKV 92
Db 11 PGFCSPRYGLAILLHFCNIAIMAQRCVCLNLTVMV-----NNTGSPHLSNESVVM 62
QY 93 LHNQTKKYRWDATOGWILGSFFYGIITQIPGGYVASRSRGKLLIGFGIFATAFTLF 152
Db 63 LDNVKNPVHSWSDIOGLLSSVFFGVMVVOAPVGLSGIYPMKRIIGSSLSFLUSSLSLL 122
QY 153 TPLAADFGVALVALRALEGEGVTYPAMHAMWSSWAPPLERSKLLISYAGAQLGTVV 212
Db 123 IPPAAQVGAALVIVCRVLOGIAGVSTGQHEIWKVWAPPLERGLTSMTSGFVMPFI 182
QY 213 SLPLSGVICYMNWTVYFFFGIVGIWFIWILWICLVSDTPETHKTIPTPYEKEYILSLKN 272
Db 243 QASGGQSLPIKAMKSLPLWAILNSFAFINSLSLVTPFTFSTVLHVNVRENGLLS 302
QY 332 AVPYLGCWLCMILSGQAADNLRARNFSLWVRVRSGLIGMIGPAIFLVAAGFIGDYSL 391
Db 303 SLPYLLAYICGILAGQMSDFELTKRIFSVTVRKLTTLGSCFPCVIFIMCLLYLSNFYS 362
QY 392 AVAFLITSTLGGFCSSGFSINHLDIAPSYAGILLITNTFATIPGMIGPIARSUTPEN 451
Db 363 TVIFLTLANSTLSFSCGQLINALDIAPRYGFLKAVTALIGMFGLLISSTLAGLINQD 422
QY 452 TIGEWOTVFCIAAINVFGAIFTFLEAKGEVQNW 486
Db 423 PEYAWHKISFLMAGINVTCLVFFLEFAKEIQDWA 457

RESULT 5
NPTL_RAT
ID NPTL_RAT STANDARD; PRT; 465 AA.
AC Q62795;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Renal sodium-dependent phosphate transport protein 1 (Sodium/phosphate
DE cotransporter 1) (Na+/Pi cotransporter 1) (Renal sodium-phosphate
DE transport protein 1) (Renal Na(+)-dependent phosphate cotransporter
DE 1).
GN SLG17A1 OR NPTL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97021433; PubMed=8867793;
RA Li H., Xie Z.;
RT "Molecular cloning of two rat Na+/Pi cotransporters: evidence for
RT differential tissue expression of transcripts.";
RL Cell. Mol. Biol. Res. 41:451-460(1995).
```

```
CC CC -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: U28504; AAC52487.1; -
CC CC InterPro: IPR003662; sub_transporter.
CC CC Pfam: PF00803; sugar_cr; 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 465 AA; 51350 MW; 28BBBF8DC5C0AC52 CRC64;

Query Match 30.5%; Score 797.5; DB 1; Length 465;
Best Local Similarity 35.8%; Pred. No. 4.7e-50;
Matches 163; Conservative 94; Mismatches 189; Indels 9; Gaps 2;

QY 33 PVCSARYNLAFLSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRTSYECABHSAPIKV 92
Db 11 PGFCSPRYGLAILLHFCNIVIMAQRCVCLNLTVMV-----NKEPPLSNKSAEM 62
QY 93 LHNQTKKYRWDATOGWILGSFFYGIITQIPGGYVASRSRGKLLIGFGIFATAFTLF 152
Db 63 LDNVKNPVHSWSDIOGLLSSVFFGVMVVOAPVGLSGIYPMKRIIGSSLSFLUSSLSLL 122
QY 153 TPLAADFGVALVALRALEGEGVTYPAMHAMWSSWAPPLERSKLLISYAGAQLGTVV 212
Db 123 IPPAAQVGAALVIVCRVLOGIAGVSTGQHEIWKVWAPPLERGLTSMTSGFVMPFI 182
QY 213 SLPLSGVICYMNWTVYFFFGIVGIWFIWILWICLVSDTPETHKTIPTPYEKEYILSLKN 272
Db 183 ALLVSGFICDLGNPMVYFIVGIVCVLSLFWFILLFDDPNHPYMSSEKDYITSLMQ 242
QY 273 QL-SSQSVPTWPMKSLPLWAIIVAHFSYNWTFYTLTLPTPYMKEVLRNFNQENGFLS 331
Db 243 QVHSGRQSLPIKAMKSLPLWAILNSFAFINSNLLVTPFTFSTVLHVNVRENGLLS 302
QY 332 AVPYLGCWLCMILSGQAADNLRARNFSLWVRVRSGLIGMIGPAIFLVAAGFIGDYSL 391
Db 303 SLPYLLAYICGIVAGQMSDFLSRKIFSVVAVRKLTTLGFCPVIFVVCVLLYLSNFYS 362
QY 392 AVAFLITSTLGGFCSSGFSINHLDIAPSYAGILLITNTFATIPGMIGPIARSUTPEN 451
Db 363 TVIFLTLANSTLSFSCGQLINALDIAPRYGFLKAVTALIGFGLISSTLAGLINQD 422
QY 452 TIGEWOTVFCIAAINVFGAIFTFLEAKGEVQNW 486
Db 423 PEYAWHKNEFLMAGINVTCLAFVLLFAKGIQDWA 457

RESULT 6
NPTL_HUMAN
ID NPTL_HUMAN STANDARD; PRT; 465 AA.
```

Q14916; Q13783;
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate
 cotransporter 1) (Na(+)/PI cotransporter 1) (Renal sodium-phosphate
 transport protein 1) (Renal Na(+)-dependent phosphate cotransporter 1)
 (NA/PI-4).
 OS Homo sapiens (Human).
 GN SLC12A1 OR NPT1.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=94117004; PubMed=8288239;
 RA Chong S.S., Kristjansson K., Zoghbi H.Y., Hughes M.R.;
 RT "Molecular cloning of the cDNA encoding a human renal sodium
 phosphate transport protein and its assignment to chromosome
 6p21.3-p23.";
 RL Genomics 18:355-359(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney cortex;
 RX MEDLINE=95126933; PubMed=7826357;
 RA Miyamoto K.-I., Tatsumi S., Sonoda T., Yamamoto H., Minami H.,
 RA Taketani Y., Takeda E.;
 RT "Cloning and functional expression of a Na(+)-dependent phosphate co-
 transporter from human kidney: cDNA cloning and functional
 expression.";
 RL Biochem. J. 305:81-85(1995).
 CC -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
 CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
 CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY CORTEX, LIVER AND BRAIN
 CC BUT NOT IN OTHER TISSUES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X71355; CAA50490.1; ALT_INIT.
 DR EMBL; D28532; BAA05888.1; -;
 DR MIM; 182308; -;
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
 KW TRANSMEM 79 99
 FT TRANSMEM 117 137
 FT TRANSMEM 176 196
 FT TRANSMEM 198 218
 FT TRANSMEM 255 275
 FT TRANSMEM 299 319
 FT TRANSMEM 337 357
 FT TRANSMEM 363 383
 FT TRANSMEM 399 419
 FT TRANSMEM 431 451
 FT TRANSMEM 439 39
 FT CARBOHYD 47 47
 FT CARBOHYD 56 56
 FT CARBOHYD 35 36
 FT CONFLICT 77 77
 FT CONFLICT 207 207
 FT CONFLICT 231 231
 FT CONFLICT 231 231
 SQ SEQUENCE 465 AA; 50884 MW; 687A19F245C4ED16 CRC64;

Query Match 30.2%; Score 790.5; DB 1; Length 465;
 Best Local Similarity 37.7%; Pred. No. 1.5e-49;
 Matches 177; Conservative 78; Mismatches 177; Indels 37; Gaps 6;
 QY 33 PVCCARYNLAFLSFFGVLYSLRVNLSVALVDVDS-----NTTAKDNRTSYCAE 85
 DB 11 PGFCFRYGLSLVHCCNVIIAQRACNLNLTWVMVNVNDPDRGLPNTSTK----- 60
 QY 86 HSAPIKVLHNTQKKYRWDATQGWILGSFFGYITITQPGYVASRSGKLLGLGIFA 145
 DB 61 -----KLLDNKNPMYNMSPDIQIILSTSYGVIIIVPVGYFGIYSTKKMIGFALCL 115
 QY 146 TAIFTLFTPLAADFGVALVALRALEGVGYTPAMHAMSSWAPPLERSKLLSISYAG 205
 DB 116 SSVLSLLIPPAAGIGVAVVVVCRVQGAAGIVATAQAEIYVVKVAPPLERGLTSMSTG 175
 QY 206 AQLGTVVSLPLSGVICYNNWTVYVFFGIVGIIIFILWICLVSDTPETHKTIITYEKEY 265
 DB 176 FLTGPFIVLLVTGVICESLGNWPMVYIFGACGCACVCLLWFLVFDPKDHPICISSEKEY 235
 QY 266 ILSLKNOL-SSQKVPWIPMLKSLPLWAIIVAHFSYNWTFYTLTLTLPYMKVLRNI 324
 DB 236 ITSSLVQQVSSRSQSLPIKAILKSLPVWALSIGSTFFWSHNMITYTPMFINSMLHVN 295
 QY 325 QENGLSAPVYLGWCLMILSQANLNRARNFNTLWVRVSLGIMIGPAIFLVAAGF 384
 DB 296 KENGLSLLPYLFAWICGNLQGLSDFELTRNLSVIAVRKLTAAAGFLPAIFGVCLPY 355
 QY 385 TCDYSLAVAFITSTTGGFCSSGFSNHLDIASVAGILGINTTATIGMIGPIIA 444
 DB 356 LSTFYSIVIFILAGATGSCFGLGVFINGLDIAPRYGFI----KACSTLGMIGGLIA 411
 QY 445 RLST-----PENTIGEWOTVFCIAAIVNFGAIFFTLFAKEGVQNW 486
 DB 412 SLTGLILKQDPESA---WFKTFILMAINVTGLIFYLIVATAEIQDWA 457
 RESULT 7
 YRT3_CAEEL STANDARD; PRT; 544 AA.
 ID YRT3_CAEEL
 AC Q10046;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 60.5 kDa protein T07A5.3 in chromosome III.
 GN T07A5.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Buck D.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z48055; CAA88134.1; -;
 DR WormPep; T07A5.3; CE01648.
 KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
 KW Sodium transport.
 FT TRANSMEM 50 70
 FT TRANSMEM 105 125
 FT TRANSMEM 128 148

```
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 194 214 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
SQ SEQUENCE 544 AA; 60543 MW; 4615B3C27D9F86CC CRC64;

Query Match 29.3%; Score 766.5; DB 1; Length 544;
Best Local Similarity 35.6%; Pred. No. 9.2e-48;
Matches 164; Conservative 90; Mismatches 166; Indels 41; Gaps 9;

QY 39 RYNLAFLSFFGFFVLYSLRVNLSVALVDMVDSNTTAKNRTSYECAHSAPIKVLINQIG 98
DB 47 RMOIATLAHFGFAISGIRSNFGVAKNRNVNFTDAYGE-----VHE--- 88
QY 99 KKYRDAETQGTILGTFYGYIITIPGGVVASRSGKLLGFGIFATIFTLETPLAAD 158
DB 89 REFLWTGAEVGMESFFGYAASQIPAGVLAAKFAPNKFIMLGILVASFMIILSAISFN 148
QY 159 FGVA---LVALRALEGLGEGVTYPAMHAMWSSWAPPLERSKILSISYAGAQIGTVVSLP 215
DB 149 FHPYTDIFVMVQVQGLGALGVLPAMHGVWKPWAPPLERSKLATTAFTGSSGVVMTGLP 208
QY 216 LSGVICYNNWTVYFFFGVIGIINFLICLVSDPETHKTIPTPYEKEYILS-----SL 270
DB 209 ASAYLSHFSWSPPFVFGVGLIINSLIMWYVSSHSPETHGYISDDKKQVTEKIGDVAV 268
QY 271 KNOLSKQSPVPTPMLKSLPLWAIVVAHFSYNNVFTYLLTLLPTMYKEVLRNIOENGFL 330
DB 269 KNM--SLTLPWRDMTSSAVWAIITCTCRSGWGFLLNQLTYMKDVLHDIKNSGFI 326
QY 331 SAVPYLGCWMLCMLSQQAADNLRARNWFTLWVRRVFSLIGMIGPAIFLVAAGFIGC--- 387
DB 327 SIPPQFGMCIVTLATQGLCDYLRSSGKMSSTEAVRKSVENTFG-----FTVEMMLCCLAF 380
QY 388 --DYSLAVAFLTISTTLGFCSSGFSINHLDTAPSAGILLGITNTFATIPGMIGPIIAR 445
DB 381 VRDPVIAVCLVITACVSGSVLSGFWNVNHFDAIPRYAPILMTGIANGLAGVAG--VGMVVTN 439
QY 446 SLTPENTIGETQVFCIAAANVFGAIFFTFLAKGVQWNA 486
DB 440 TTYIQNPDG--WKWVFLAMADIFGVIFFLIFAKGDVLPWA 479

RESULT 8
NPT3 HUMAN
ID NPT3 HUMAN STANDARD; PRT; 436 AA.
AC O00624;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sodium-dependent phosphate transport protein 3 (Sodium/phosphate
DE cotransporter 3) (Na(+)/PI cotransporter 3).
GN SLC17A3 OR NPT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=9149941;
RA Ruddy D.A., Kronmal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA Domingo R. Jr., Meyer N.C., Irrinki A., McClelland E.E., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.
RT "A 1.1-Mb transcript map of the hereditary hemochromatosis locus.";
RC Genome Res. 7:441-456(1997).
CL -!- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
```

```
CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U91328; AAB82085.1; -
DR EMBL; U90544; AAB53422.1; -
KW Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
KW Sodium transport.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 366 406 POTENTIAL.
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 25.7%; Score 672; DB 1; Length 436;
Best Local Similarity 38.9%; Pred. No. 4.2e-41;
Matches 139; Conservative 69; Mismatches 133; Indels 16; Gaps 4;

QY 27 PRAEPAPVCCSARYNLAFSLFFGFFVLYSLRVNLSVALVDMVDSNTTAKNRTSYECAH 86
DB 5 PATRKGPDPCSLRYGLALIMHFSNFTMTITQRVSLTALIMV--NTQOQGLSN---AST 59
QY 87 SAPIKVLHNOT-----GKKYRWDAETQGTILGTFYGYIITIPGGVVASRSGGK 136
DB 60 EGPVADAFNNSSISKEFDTKASVQWSPETQGIFFSSYNGIILLIIFSGVLAGIFGAK 119
QY 137 LLLGFGIFATIFTFTPLAADFGVALVALRALEGLGEGVTYPAMHAMWSSWAPPLERS 196
DB 120 KMLGAGLLISLLTFTPLAADFGVILVIMVTVQGMAGMAWTGQFTIWKAWAPPLERS 179
QY 197 KLLSISYAGAQIGTVVSLPLSGVICYNNWTVYFFFGVIGIINFLICLVSDPETHK 256
DB 180 KLTITAGSGSAFGSFTILCVGGLISQALSQWPFIFYIFGSGTCVCCLLWFTVIYDDPMHP 239
QY 257 TITPYEKEYILSLKNQLSSQ--KSPVWIPMLKSLPLWAIVVAHFSYNNVFTYLLTLPY 315
DB 240 CTSVREKEHILSLAQOQSPSPGRAPVPIKAWVTCLPULWALFGLFFSHFWLCTIILTYPIY 299
QY 316 MKEVLRNFNIQENGFLSAVPLGCLWMLCSGQAADNLRARNWFTLWVRRVFSLIGM 372
DB 300 ISTLLHVNIRDSGVLSLSPFIAAASCTILGGQLADFLLSRNLRLTLTKRKLSSLDL 356

RESULT 9
NPT4 HUMAN
ID NPT4 HUMAN STANDARD; PRT; 401 AA.
AC O00476;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sodium-dependent phosphate transport protein 4 (Sodium/phosphate
DE cotransporter 4) (Na(+)/PI cotransporter 4).
GN SLC17A4 OR NPT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```



```
Db 45 SRLHLVLLCMIGFCTTMRHLFALTMCMVNSTALAVENEIKLAGNSNVSEISIEI 104
QY 93 -----LHNOTGKK-----YRMDAETQGWILGSFFGYIITQIPGGYASRSG 134
Db 105 NLGNSQCGLMDEGQKKVVVDYGGELVNSYEQNLIFSGTFWGSLLTVLPSMFFIERES 164
QY 135 GKLLGFGIFATFTLTP-LAADFGVALVALRALEGLGEGVTVYPAMHAMSSWAPPL 193
Db 165 PRHVLOISVALYTLVTVITPFLATHEGYEVLARIGMGLGEGFVPTNNAIGNWFPSS 224
QY 194 ERSKLLISYAGAOGLTVVSLPLSGVICYY-MNWTVVYFFGVLGVLWILWILCLVSDTP 252
Db 225 EKSTALISITLGNQISAAGSPMAAACSDLGWPATFYFAGIFATGWSLWILWFFTASSHP 284
QY 253 ETHKTTTPYEKEYILSSLNQ-----SSQKSVPMPLKSLPLWAIIVVAHFSYNWTFYTL 309
Db 285 AKVKMTKEKEYILLANVKKVHKSEKTSIPYSKILTSPAFGLQLOCHFEVFN-LFWTLF 343
QY 310 TL-LPTMYKEVLRFNQENGFLSADPYLGCWLCMLISGQAADNLRARNWFTLWVRVFS 368
Db 344 QIYLPSEYFEVLHGLVIANGTETAIPNIFNMIFKVVWVGIGIDKLENKILSNTKAVKVSH 403
QY 369 LIGMIGPAIFLV-AAGFIGCDYSLA--VAFLTISTTLGFCSSGFSINHLDTAPSAGIL 425
Db 404 GVASFGSSSLILLAFVDCSNPTTGLIFFCLMYSSMGTFV-SGFTSLLSLAPQYATM 462
QY 426 LGTINTFATIPGMIGPIIARSITPENTIGEWOTVFCIAAANVGAIFFTLFAKGEVQNW 485
Db 463 SALSMEFVAMIGRLTTPAVMSMERKDGTAEFQWONIFGCSLAHIFSGSIFLLFGSELQDW 522
QY 486 A-ISDHQ 491
Db 523 AKVEDDQ 529

RESULT 11
GUDP_BACSU STANDARD; PRT; 455 AA.
AC P42237;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable glucarate transporter (D-glucarate permease).
GN GUDP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219079; PubMed=7704254;
RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
RT "Determination of a 21548 bp nucleotide sequence around the 24
RT degrees region of the Bacillus subtilis chromosome.";
RL Microbiology 141:269-275(1995).
CC -!- FUNCTION: UPTAKE OF D-GLUCARATE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; D30808; BAA06469.1; -
DR EMBL; Z99105; CAB12042.1; -
DR Subtilist; BG11160; gudp.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; suga_tr; 1.
```

```
Transmembrane; Transport; Complete proteome.
KW FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 253 273 POTENTIAL.
FT TRANSMEM 289 309 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 414 434 POTENTIAL.
SQ SEQUENCE 455 AA; 49254 MW; A0DFD23F569CC7B8 CRC64;

Query Match 14.3%; Score 373; DB 1; Length 455;
Best Local Similarity 25.5%; Pred. No. 1.le-19;
Matches 120; Conservative 80; Mismatches 211; Indels 60; Gaps 13;

QY 31 PAPVCCSARYNLAFSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRTSYECAHSAPI 90
Db 10 PAGKKTSVWFIVFMLFLVTSINYADRALSI-----TGDSVQHDLC 52
QY 91 KVLHNOTGKKYRDAAETQGWILGSFFGYIITQIPGGYASRSGKLLGFGIFATAFT 150
Db 53 -----DSVAMGYVFSAGWAVYIQLPGWLLDRFGSKTIIALSIFWWSFT 99
QY 151 LFTPLAADFGVCA-----LVALRALSGLEGVTVYPAMHAMSSWAPPLSKLLSISYAGA 206
Db 100 LLQGAIGFFSAGTAIILLFALRFLVGLSEAPSPGNGRVYASWFPSSERGTASAFNSAQ 159
QY 207 QLGTVVSLPLSGVICYMNMTVYVFFGIVGTIITWFLTWICLVSDTPETHKTTTPYEKEYI 266
Db 160 YFAIVIFPLMGWLTHSFCWHSVYVVMGVIAGILLAVIMLKTVE-PKKHPKVNEALAVI 218
QY 267 -----LSSLKNQLSQKSVPMPLKSL-----PLWAIIVVAHFSYNWTFYTLTLLPTMYK 317
Db 219 EOGGLISMDDSKSQETESKWPYIKQLLTNRMLIGVYIAQYCIITLTLYFFLTWFPVYLV 278
QY 318 EVLRNENIQENGFLSADPYLGCWLCMLISGQAADNLRARNWFTLWVRVFSLIGMIGPAI 377
Db 279 QARGMSILEAGFVASLPCALCGFAGVGLGVSDILLKKGR-SLTFAKVPPIIAGM----- 332
QY 378 FLVAAGFIGCDYS-----LAVAFITISTTLGFCSSGFSINHLDTAPSAGILGIGTNTFF 432
Db 333 -LLSCSMIVCNVYDTSANLWVIMSLAFCGKFGALGWAVVS-DTSPKECAGLSGGLNTFF 390
QY 433 ATIPGMIGPIIARSITPENTIGEWOTVFCIAAANVGAIFFTLFAKGEVQ 483
Db 391 GNIASITTPIIIGYIV--NATGSFNGALVFVGA-NAIAILSYLLVIGPIK 438

RESULT 12
GUDP_ECOLI STANDARD; PRT; 450 AA.
AC Q46916;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable glucarate transporter (D-glucarate permease).
GN GUDP OR B2789.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
```


CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26185; BAA0188.1; -;
CC EMBL; Z99124; CAB16094.1; -;
CC Subtilist; BG10016; ybO.
CC InterPro; IPR003662; sub_transporter.
CC Pfam; PF00083; sugar_tr_1;
KW Hypothetical protein; Transmembrane; Transport; Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
SQ SEQUENCE 435 AA; 48248 MW; 9630914D9B606208 CRC64;

Query Match 12.0%; Score 314; DB 1; Length 435;
Best Local Similarity 25.3%; Pred No. 1.8e-15;
Matches 116; Conservative 71; Mismatches 192; Indels 80; Gaps 18;
QY 49 GFFVL-YSLRVNLSVALVDMVDSNTAKDNRSTYECAEHSAPIKVLHNQTKGKYRWDAET 107
DB 29 GIILNFDRAISVAAPAIQDS-----FHLTATE 58
QY 108 QGWILGSFFYGYITQIPGGVVASRSGKLLLGFGIFATAITFTPLAADFGVALVAL 167
DB 59 LGIVFSIYTYSTLMQLPVGSLDRFGVAVTRVGMTIWSFLTIL--LAFLLQKLLLYLF 116
QY 168 RALEGLSGGVTPAMHAMWSSWAPLERSKLLSISYACQALGTVVSLPSGVICYMNT 227
DB 117 RFLGLTSASAPPAASKATALWFPSPERGLANSUFDAAKFSNVIGAPLVAFVLTTFDWR 176
QY 228 YVYFFGIVGIWFIWILICLVSDTPETHKTIPTVEKEYILSSKLNQSSQKSVPMI--PM 285
DB 177 VAFLTIGCINVL-ETIFWQYQYQPERHKRISKSELNVI---QKHNAITTEQIPYKTGPL 232
QY 286 LKSL-----PLMAIVAHFSYNWTFYTLTLPTVMKEVLRNIQENGFLSAVPLGWL 341
DB 233 LKLFTRKRWGLMIGETGYTETNLLITWLPFFKHTYGMDSGLFTAVP---WLI 288
QY 342 MILSQADNLRAW-----NFTLWVRVFLSMIGMGAIFLYAAGFIGDYS 393
DB 289 STISGIVGG-----WLVDYFIKKGYPNKYRTVIIVGMSFGFFFL---GSI-LTNNTY 340
QY 394 AFLTISTTLGGFCSS---GFSINHLDIAP----SYAGILGLITNTFTATIPGMIGPIARS 446
DB 341 AIIICISIGLAGISATAPVGNIS-AELAPIGSVMSLMSWNLAN-----NLFGGITAA 393
QY 447 LTPE--NTIGEMQVFCIAAAINVFGAIFFTFLFAKGEVQ 483
DB 394 LTGYLFDVTSFTLSFLVAGFVLLGLVFEY-VFVLGDVK 431

Search completed: July 15, 2002, 08:25:27
Job time: 192 sec

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB ID	Description
1	2617	100.0	495	6 Q9WZD1	Q9WZD1 ovis aries
2	2329	89.0	495	4 Q9UGH0	Q9UGH0 homo sapien
3	2329	89.0	536	4 Q9NR42	Q9NR42 homo sapien
4	991.5	37.9	559	5 Q9VYG7	Q9VYG7 drosophila
5	985.5	37.7	502	5 Q9VDM0	Q9VDM0 drosophila
6	965	36.9	582	11 Q9JTL2	Q9JTL2 rattus norv
7	965	36.9	582	11 Q9Z0B7	Q9Z0B7 mus musculu
8	962.5	36.8	582	4 Q9PZ08	Q9PZ08 homo sapien
9	941	36.0	529	5 Q9V7S5	Q9V7S5 drosophila
10	936	35.8	586	5 Q23514	Q23514 caenorhabdi
11	932	35.6	560	4 Q9P2U7	Q9P2U7 homo sapien
12	927	35.4	560	11 Q62634	Q62634 rattus norv
13	926	35.4	483	5 Q61369	Q61369 drosophila
14	884	33.8	466	4 Q96LH1	Q96LH1 homo sapien
15	874.5	33.4	497	4 Q9V2C5	Q9V2C5 homo sapien
16	859.5	32.8	481	5 Q9VJW8	Q9VJW8 drosophila

Db	121	ITQIPGGVVASRSGKLLLGFGIFATAFTLTFTPLAADFGVALVALRLEGSGVTYP	180
Qy	181	AMHAMWSWAPPLERSKLISISYAGAQLGTVVSLPSGVICVYMMNWTYVFFFGIVGIW	240
Db	181	AMHAMWSWAPPLERSKLLSISYAGAQLGTVVSLPSGVICVYMMNWTYVFFFGIVGIW	240
Qy	241	FIWTICLVSDTPETHTKTIPTPYEKEYILSSLKNOLSSQKSVWPIMKLSPLWAIIVVAHFS	300
Db	241	FIWTICLVSDTPETHTKTIPTPYEKEYILSSLKNOLSSQKSVWPIMKLSPLWAIIVVAHFS	300
Qy	301	YNNTEYTLTLTLLPTYMKVEVLNRNIQENGFSLSAVPLYGCWLCMLISGQADNLRARNEST	360
Db	301	YNNTEYTLTLTLLPTYMKVEVLNRNIQENGFSLSAVPLYGCWLCMLISGQADNLRARNEST	360
Qy	361	LWVRVFSLIGMIGPAIFLVAAGFTGCDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS	420
Db	361	LWVRVFSLIGMIGPAIFLVAAGFTGCDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS	420
Qy	421	YAGILLGTTNTFATIPGMIGPIIARSLTPENTIGEWQTVFCIAAINVGAIFFTLFAGK	480
Db	421	YAGILLGTTNTFATIPGMIGPIIARSLTPENTIGEWQTVFCIAAINVGAIFFTLFAGK	480
Qy	481	EYQNWAISDHQHRN 495	
Db	481	EYQNWAISDHQHRN 495	
RESULT 2			
Q9UGH0 PRELIMINARY; PRT; 495 AA.			
AC	Q9UGH0		
DT	01-MAY-2000 (TReMBLrel. 13, Created)		
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)		
DE	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	SIALIN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20047778; PubMed=10581036;		
RA	Verheijen F.W., Verbeek E., Aula N., Beereus C.E.M.T., Havelaar M.C.,		
RA	Joesse M., Peltonen L., Aula P., Galjaard H., Van der Spek P.J.V.D.,		
RA	Mancini G.M.S.;		
RT	"A new gene, encoding an anion transporter, is mutated in sialic acid		
RT	storage diseases.";		
RL	Nat. Genet. 23:462-465(1999).		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.		
DR	EMBL: AJ387747; CAB62540.1; -		
DR	InterPro: IPR003662; sub_transporter.		
DR	Pfam: PF00083; sugar_tr; 1.		
KW	Transmembrane.		
SQ	SEQUENCE 495 AA; 54639 MW; 5C6C154B3E93A19E CRC64;		
Query Match 89.0%; Score 2329; DB 4; Length 495;			
Best Local Similarity 86.3%; Pred. No. 8.5e-175;			
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;			
Qy	1	MKSPVSDLPADSGEGSDRTPLIQAPRAEPAPVCCSARYNLAFISFTGFFVLYSLRNL	60
Db	1	MKSPVRDLARNDGEESTDRTPLLPAPRAEAPVCCSARYNLALIAFFGFFIVVALRNL	60
Qy	61	SVALVDVDSNTAKRDNITSYCEAHSAPIKVLHNQTKKRWDAETOGWILGSFFGYI	120
Db	61	SVALVDVDSNTLTEDNRTSKACPEHSAPIKVHHNQTKKQWDAETOGWILGSFFGYI	120
Qy	121	ITQIPGGVVASRSGKLLLGFGIFATAFTLTFTPLAADFGVALVALRLEGSGVTYP	180
Db	121	ITQIPGGVVASRSGKLLLGFGILGTAVLTFTPLAADLGVGLIVLRLEGSGVTYP	180

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PUTATIVE INORGANIC PHOSPHATE COTRANSPORTER (PICOT PROTEIN) (CG8098
 DE PROTEIN).
 GN PICOT OR CG8098.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Cocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champ M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ihegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh K.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RP [2]
 RN SEQUENCE OF 449-529 FROM N.A.
 RC STRAIN=CANTON-S;
 RA D Lage J.-L.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
 CC -1- SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL; AE003806; AAF57968.1; -;
 DR EMBL; AE003806; AAF57969.1; -;
 DR EMBL; AF022713; AAD09148.1; -;
 DR FlyBase; FBgn0024315; Picot.
 DR InterPro; IPR003362; sub.transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 KW Alternative splicing; Transport; Transmembrane.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 202 222 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 338 358 POTENTIAL.
 FT TRANSMEM 429 449 POTENTIAL.

FT TRANSMEM 466 486 POTENTIAL.
 FT VARSPLIC 1 35 MPERSSLNHRHRRGHVLRVWNRNHSLEQOPOR -> MS
 FT ASKEAICGSGTEKDLKPALG (IN SHORT ISOFORM).
 SQ SEQUENCE 529 AA; 58372 MW; 07B89A52D5081EFB CRC64;
 Query Match 36.0%; Score 941; DB 5; Length 529;
 Best Local Similarity 40.8%; Pred No 8 6e-66;
 Matches 193; Conservative 85; Mismatches 177; Indels 18; Gaps 8;
 QY 23 LQAPRAEPAPVCCSARYNLAFLSFFGVFLYSLRVNLSVALVDMVDSNTAKDNRTSY- 81
 DB LEQOPQR-----CFATRYFVTEFLGMAAYVMRTNMSVAIVMNHATAIKSGEAEYD 83
 QY 82 -ECAHSAPIKVLHNOTGKKYKWDATQGWILGSPFYGVITQIPGYYVASRGGKLLIG 140
 DB DECGDRDIPID--DSQDG-EFAWSSALQGYLSSFFYGVITQIPGILAKKYGSURFLG 140
 QY 141 FGIFATAFTLTPLAA-DFGVGALVALRLEGVGGVTPYPAHAMWSSWAPLERSKLL 199
 DB YGLINSVFAFLVPVAARGGVWGLCAVRFIOGLGEGPIVPCHTAMAKWIPNERSMG 200
 QY 200 SISIYAGAQGTGVVSLPLSGVICVY---MNVTVYFVGIVGIWFIWICLVSDTPEHK 256
 DB AAVYAGAQGTIISMLPSGLLAIEYFGDGWPSIFVYFVGIVGTWSTAFILFVHEDPSSHP 260
 QY 257 TITPYKEVILSL-KNOILSSQKSVPIWMLKSLPLWAIWAHFSYNWTFYLLTLLPTY 315
 DB TIDEREKKYINDSLWGTGVVSPPIPKAIISLPYAILFAHMGHNYGYETLMTLPY 320
 QY 316 MKEVLRNFOENGLFSAPVPLGVCMLCSGQAADNLARNWPNFSTLWVRVFSLIGMIGP 375
 DB MKQVLRFSKLSGLLSLPLWLFESMFSIVADVDMWISKRFSTATKRLNSIGQYGP 380
 QY 376 AIFLVAAGFIGDCYSLAVAPLFTITTLGGFCSSGFSINHLDIAPSAGVAGILLGTNTFAT 435
 DB GVALIAASYTGCDRALTLAILTGVGLNGGIYSGGFNHLDLTPRPAGFLMSITNGSAML 440
 QY 436 PGMGIIPIARSITPENT---IGEWQVFCIAAINVGAIFFTLPAKGEVQNW 485
 DB AGLIAPIAAGHLISDPSKPMGQWQIVFFIAAFVYIICGTFYIFGSGGROIW 493
 RESULT 10
 ID Q23514 PRELIMINARY; PRT; 586 AA.
 AC Q23514;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOTHETICAL 65.0 KDA PROTEIN.
 GN ZK54.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R., Gattung S., Le T.T.;
 RT "The sequence of C. elegans cosmid ZK54.";
 RN Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;

```
RA Waterston R.;
RT "Direct submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBDJ databases.
KW EMBL; U58737; AAK39396.1; -.
KW Hypothetical protein.
SQ SEQUENCE 586 AA; 65045 MW; 95D52B7F7FA7DD0F CRC64;

Query Match 35.8%; Score 936; DB 5; Length 586;
Best Local Similarity 42.0%; Pred. No. 2.4e-65;
Matches 192; Conservative 82; Mismatches 155; Indels 28; Gaps 8;

QY 39 RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTAKDNKTSYECAEHSAPIKVLNQTG 98
Db 145 RHVVAILALGFANFIYAMRANLSIAIVEMT-SGTERKVNCT-----LH-VLG 190
QY 99 KKYRWDATQGTLSGFFGYIITQIPGGVVASRSRGKLLGFGIFATAFTLTPLAAD 158
Db 191 DFNWTPMTQGVVLSFFGYIVSQLPGGYLATYHGAKTFFAGTGTAVFTLTTPPEAR 250
QY 159 FGVALVALRALGEGGVTPYPAHMAWSSWAPPLERSKILLSYAGAOIGTVVSLPLSG 218
Db 251 MGYGMLVFARFMGLEGVTPYPAHVIWSRWAPPMQTKLATFAFSSYFGTVVAMPLSA 310
QY 219 VICYYMNWTVVFFGVIGLWILMICLVSDPTPEKHTITPYEKEYIILSSLNQSLSSQK 278
Db 311 YLGEHFGWPMIEFFGALGVIMCWVYKTVHDPEDDPDKISTSE----LALLQDAVSQN 366
QY 279 S--VPMTPLKSLPLAAIVVAHESYNWTFVTLTLTPYMKVLRNIQENGELSAVPYL 336
Db 367 HYIVPQAQLRSRPWAWIVAHSAQNLGFGYIMTLNPKMKLDIAGYNVEKAGIASLPHY 426
QY 337 GCWLCMLISQQAADNLRARWNFTLWVRVFLSGLIGMIGPAIFLVAAGFTGCDYS---LAV 393
Db 427 LMGFOITGGQLCDYLRKDKHYDTLFRKMACALGFIGQSVEL----FLWMTTSNLLV 482
QY 394 AFTISTTTLGGFCSSGFSINHLDIAPSYAGILGINTNTATIPGMIGPIIARSILTPNTI 453
Db 483 LFFSISITGLGICWCWFSVNLHDLAPQYAGHLMATSNATIPGIFPLLVGAIVQNGTI 542
QY 454 GEMQTVFCIAAANVFGAIFFTFLFAKGEVQNWAISDH 490
Db 543 GEMNVIMYIIISAYLIGAAIFKFAFDATLQPPWA-AEH 578

RESULT 11
Q9P207 PRELIMINARY; PRT; 560 AA.
AC Q9P207
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE BRAIN-SPECIFIC NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
GN ENPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Aihara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T.,
RA Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
RT "Molecular cloning of a novel brain-type Na+-dependent inorganic
RT phosphate cotransporter."
RL J. Neurochem. 0:0-0(2000).
DR EMBL; AB032436; BAA92875.1; -.
SQ SEQUENCE 560 AA; 61613 MW; C88DAFB34B6E45B6 CRC64;

Query Match 35.6%; Score 932; DB 4; Length 560;
Best Local Similarity 39.5%; Pred. No. 4.7e-65;
Matches 191; Conservative 92; Mismatches 171; Indels 30; Gaps 6;

RA Waterston R.;
RT "Direct submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBDJ databases.
KW EMBL; U58737; AAK39396.1; -.
KW Hypothetical protein.
SQ SEQUENCE 586 AA; 65045 MW; 95D52B7F7FA7DD0F CRC64;

Query Match 35.8%; Score 936; DB 5; Length 586;
Best Local Similarity 42.0%; Pred. No. 2.4e-65;
Matches 192; Conservative 82; Mismatches 155; Indels 28; Gaps 8;

QY 39 RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTAKDNKTSYECAEHSAPIKVLNQTG 98
Db 145 RHVVAILALGFANFIYAMRANLSIAIVEMT-SGTERKVNCT-----LH-VLG 190
QY 99 KKYRWDATQGTLSGFFGYIITQIPGGVVASRSRGKLLGFGIFATAFTLTPLAAD 158
Db 191 DFNWTPMTQGVVLSFFGYIVSQLPGGYLATYHGAKTFFAGTGTAVFTLTTPPEAR 250
QY 159 FGVALVALRALGEGGVTPYPAHMAWSSWAPPLERSKILLSYAGAOIGTVVSLPLSG 218
Db 251 MGYGMLVFARFMGLEGVTPYPAHVIWSRWAPPMQTKLATFAFSSYFGTVVAMPLSA 310
QY 219 VICYYMNWTVVFFGVIGLWILMICLVSDPTPEKHTITPYEKEYIILSSLNQSLSSQK 278
Db 311 YLGEHFGWPMIEFFGALGVIMCWVYKTVHDPEDDPDKISTSE----LALLQDAVSQN 366
QY 279 S--VPMTPLKSLPLAAIVVAHESYNWTFVTLTLTPYMKVLRNIQENGELSAVPYL 336
Db 367 HYIVPQAQLRSRPWAWIVAHSAQNLGFGYIMTLNPKMKLDIAGYNVEKAGIASLPHY 426
QY 337 GCWLCMLISQQAADNLRARWNFTLWVRVFLSGLIGMIGPAIFLVAAGFTGCDYS---LAV 393
Db 427 LMGFOITGGQLCDYLRKDKHYDTLFRKMACALGFIGQSVEL----FLWMTTSNLLV 482
QY 394 AFTISTTTLGGFCSSGFSINHLDIAPSYAGILGINTNTATIPGMIGPIIARSILTPNTI 453
Db 483 LFFSISITGLGICWCWFSVNLHDLAPQYAGHLMATSNATIPGIFPLLVGAIVQNGTI 542
QY 454 GEMQTVFCIAAANVFGAIFFTFLFAKGEVQNWAISDH 490
Db 543 GEMNVIMYIIISAYLIGAAIFKFAFDATLQPPWA-AEH 578

RESULT 11
Q9P207 PRELIMINARY; PRT; 560 AA.
AC Q9P207
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE BRAIN-SPECIFIC NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
GN ENPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Aihara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T.,
RA Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
RT "Molecular cloning of a novel brain-type Na+-dependent inorganic
RT phosphate cotransporter."
RL J. Neurochem. 0:0-0(2000).
DR EMBL; AB032436; BAA92875.1; -.
SQ SEQUENCE 560 AA; 61613 MW; C88DAFB34B6E45B6 CRC64;

Query Match 35.4%; Score 927; DB 11; Length 560;
Best Local Similarity 39.4%; Pred. No. 1.2e-64;
Matches 184; Conservative 89; Mismatches 168; Indels 26; Gaps 4;

QY 30 EPAPVCCSA-----RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTAKDNRTSYECA 84
Db 48 DPPVWVDTCTGFLPRRYIIAIIIMSGLCISFGIRCNLGLVAIVSMVNSTTHRGHVVQKA 107
```

Qy	85	EHSAPIKVLHNQTKKKYRWDARTQGWILGSPFYGYIITQIPGGYVASSRSGGKLLIGFGIF	144
Dy	108	Q-----FNWDPETVGLIHGSPFWGIVTQIPGGFICQKFAANRVGFAIV	152
Qy	145	ATAFTLFTPLAADRGVAGALVALRLEGIGEGVTPYPMAMWSSNAPPLERSKLLISYA	204
Dy	153	ATSTLNLIPSNARVHYGCVIFVRILQGLVGGVTPACHGIWSKWAPPLERSRLATAPC	212
Qy	205	CAQLGTVYSLPLSGVICYIMMWTYVYFPGVIGIILFWILMCLVSDTPTHTKITIYPKE	264
Dy	213	GSYAGAVAMPLAGVLVQYSSWSSVYVYVSGFIWYFLWLLVSVESPALHPSISEERK	272
Qy	265	YILSLKNQLSQSV-----PWTPMLKSLPLWALVWAHFSYNNWTFYLLTLPLPYMKEV	319
Dy	273	YIEDAIGESAKLNNPVTENTPWRFFTSMPVYALIVANFCRSWTFYLLILISQPAFEEV	332
Qy	320	LRFNIQENGELSAPVYLCWCLMILSGOAAADNLRARWNFSTLWVRVFSLCMIGIPAFL	379
Dy	333	FGFELSKVGLVSLADPHLVMTIIVPIGGQIADFLSRGRHMSTTNVRKLMNCGGFGMEATLL	392
Qy	380	VAAFGIGCDYSLAVAFLLISTTLGFCSSGFSINHLDIAPSYAGILLGTINTTFATIPGI	439
Dy	393	LVVGY-SHSGKVAISFLVAVGFSGFAISGFNVNHLDIAPYASILMGISNCGVGLSGMV	451
Qy	440	GPIATARSILTPTIGEMQTVFCIAAAINVFGAIFTFLFAKEGVQWA	486
Dy	452	CPIIVGAMTKHKTREOWYVFLIASLVHYGVIFYGVFASGEKQPA	498
RESULT 13			
O61369			
ID	O61369	PRELIMINARY; PRT; 483 AA.	
AC	O61369;		
DT	01-AUG-1998	(TREMBLrel. 07, Created)	
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	PUTATIVE INORGANIC PHOSPHATE COTRANSPORTER.		
GN	PICOT.		
OS	Drosophila ananassae (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7217;		
EN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-TAI 13-1610;		
RA	DA Lage J.-L., Alland C.;		
RL	Submitted (SPP-1997) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
DR	EMBL; AF024691; AAC39088.1; -;		
DR	FlyBase; FBgn0024472; Dana\Picot.		
DR	InterPro; IPR003662; sub_transporter.		
DR	Pfam; PF00083; sugar_tr; 1.		
KW	Transport; Transmembrane.		
FT	TRANSMEM 187 207	POTENTIAL.	
FT	TRANSMEM 292 312	POTENTIAL.	
FT	TRANSMEM 383 403	POTENTIAL.	
FT	TRANSMEM 420 440	POTENTIAL.	
SQ	SEQUENCE 483 AA; 52885 MW; 6832750611B775FO CRC64;		
Query Match 35.48; Score 926; DB 5; Length 483;			
Best Local Similarity 42.38; Pred. No. 1.2e-64;			
Matches 190; Conservative 81; Mismatches 164; Indels 14; Gaps			
Qy	47	FFGFVLYSLRWNLVSVALVDMVDSNTTAKDNRTSY--ECASHSPAKVLHNQTKKKYRWD	104
Dy	3	FLGMANAVMTNNSVAIVAMV-NHTAIKSGEEYDDECGDRDIPID--DSODG-EFPWN	58
Qy	105	AETQGWILGSPFYGYIITQIPGGYVASSRSGGKLLIGFGIFATAFTLFTPLAA-DFGVGA	163
Dy	59	AALQGYILSSFPYGYIITQIPGFLAKKYGSLRFLGYGLMINSVPAFLVPVAAREGGVWG	118

[illegible]

Matches 190; Conservative 81; Mismatches 164; Indels 14; Gaps 8;
Best Local Similarity 42.3%; Pled. NO. 1.2E-04;

QY 325 QENGFLSAVYPLGCVLMLSCQADNLPRNWFSTLWVRVFSLIGMIGPAIFLVAAGF 384
Db 297 RDSGVLSSLPFTAAASCTILGQDLALLSRNLLRLITVTKLFSSGLGLLPSCICAVLPF 356
QY 385 IGCXYSLAVAFLTISTLGGFCSSGFSINHLDIAPSYAGILLGITNTFATIPGMIPIA 444
Db 357 VASSVYIITLILIPGNSLNCDSGFINTLDIAPRYASFLMGSRGFLGIAGIISSTAT 416
QY 445 RSLTPENTIGEQTVFCIAAAINVFGAIFFTLFAKGEVQNA 486
Db 417 GFLISQDFSGWRNVFLLSAAVNMGFLVYFLTFGQAEIQDWA 458

RESULT 15
Q9Y2C5 PRELIMINARY; PRT; 497 AA.
AC Q9Y2C5;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE NA/PO4 COTRANSPORTER HOMOLOG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99253143; PubMed=10319585;
RA Shibui A., Tsunoda T., Seki N., Suzuki Y., Sugane K., Sugano S.;
RT "Isolation and chromosomal mapping of a novel human gene showing
homology to Na+/PO4 cotransporter.";
RL J. Hum. Genet. 44:190-192(1999).
CC -!- SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AB020527; BAA76663.1; -
DR InterPro: IPR003662; sub-transporter.
DR Pfam: PF00083; sugar_tr.1.
KW Transmembrane.
SQ SEQUENCE 497 AA; 54055 MW; 5A5616AA52D4990D CRC64;

Query Match 33.4%; Score 874.5; DB 4; Length 497;
Best Local Similarity 37.4%; Pred. No. 1.3e-60;
Matches 185; Conservative 101; Mismatches 188; Indels 21; Gaps 7;
QY 1 MKSPYDLAPSDGEGSDRTPLLQAPRAEPAPVCCSARYNLAEISFGGFVLYSLRVNL 60
Db 7 VKATVGDIS-SDGN-----LNVAQECRKGFCSVRHGLALILQLCNFSIITQQMNL 57
QY 61 SVALYDMVDSNTAKDNRTSYECAHSAPIKVLHNQTKK-----YRWDAETQGWILG 113
Db 58 STAIAPMV--NNTAPPSPNASTERPSTDSDQYWNETLKEPKAMAPAYDWSPEIQIILS 115
QY 114 SFYGYIITQPGYVYASRGKLLIGFIFATAIFTFTPLAADFGYGVGALVALRLEGL 173
Db 116 SLNYGSFLAPIPSGVAGIFGAKYVYVAGLFSSELTFLFIPLAANAGVALLIVLRIVQGI 175
QY 174 GEGVTYPAMHAMWSSWAPPLERSKLLISYAGAOLGVVYVSLPSLGVICVYMMNTVYVFF 233
Db 176 AQVMVLTQYSIHWKAPPLERSQITTAGSGMLGSFIVLLAGLLCQTIGWPYVYIF 235
QY 234 GIVGIINFILWICLVSDPTETHKITPYEKEYILSKNQ-LSSQKSVPIPMKLSPLW 292
Db 236 GGIGCACCPFLWPLYDDPVNHPFISAGEKRYIVCSLAQQDCSPGWSLPIRAMIKSLPLW 295
QY 293 AIVVAHFSYNWTFYLLTLLPTPYMKEVLRFNQENGFLSAVPY-LGCWLCMLLSQAADN 351
Db 296 ALLVSFYCEYMLFTYIMATPYTISVILQANLRDSGILSALPFVYGC-ICILGGLLADF 354
QY 352 LRARNFSTLWVRVFSLIGMIGPAIFLVAAGFIGCDYSLAVAFLTISTTLGGFCSSGFS 411
Db 355 LLRSKILRLIIRKLTALGVLFPSPVILSLPWRSSHSMWTFLVLSAISSFCESGAL 414

QY 412 INHLDIAPSYAGILLGITNTFATIPGMIPIARSLTTPENTIGEQTVFCIAAAINVFGA 471
Db 415 VNFLDIAPRYTGFLKGLLQVFAHIAISPTAAGFFISQDSEFGWRNVFLLSAAVNISGL 474
QY 472 IFFTLLFAKGEVQNA 486
Db 475 VFYLIFGRADVQDWA 489

Search completed: July 15, 2002, 08:25:03
Job time: 168 sec

OM of: US-09-776-865-4 to: GenEmbl:* out_format : pfs
Date: Jul 15, 2002 3:40 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09776865/runat_15072002.082208.27415/app_query.fasta_1.1159
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bgsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=oct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09776865 -CGN1_15372 -NCPD=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-776-865-4
Query length: 495
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 3807.270000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_om:AF244578	+ 2617.00	4124.27	2.3e-221	2844	AF244578 Ovis aries membrane
gb_pat:AX207626	+ 2617.00	4124.27	2.3e-221	2844	AX207626 Sequence 3 from Patent
gb_pat:AX138494	+ 2329.00	3668.89	5.4e-196	2512	AX138494 Sequence 2 from Patent
gb_pr:HS387747	+ 2329.00	3668.89	5.4e-196	2512	AX387747 Homo sapiens mRNA for
gb_pat:AX207624	+ 2329.00	3667.31	6.6e-196	2330	AX207624 Sequence 1 from Patent
gb_pr:AF24577	+ 2329.00	3667.31	6.6e-196	2330	AF24577 Homo sapiens membrane
gb_pr:BC020961	+ 2329.00	3665.99	7.8e-196	3329	BC020961 Homo sapiens, solute
gb_in:AV060776	+ 985.50	1541.83	1.6e-77	1841	AY060776 Drosophila melanogast
gb_hlg:AC014246	+ 981.50	1505.63	1.7e-75	3344	AC014246 Drosophila melanogast
gb_in:AC023685	+ 981.50	1489.01	1.4e-74	167928	AC023685 Drosophila melanoga
gb_in:AC023711	+ 981.50	1488.64	1.5e-74	174163	AC023711 Drosophila melanoga
gb_in:AF003491	+ 981.50	1482.26	3.4e-74	323461	AF003491 Drosophila melanoga
gb_ro:AF324864	+ 969.00	1512.41	7.0e-76	2528	AF324864 Mus musculus vesicula
gb_ro:AF271235	+ 968.00	1506.14	1.6e-75	3382	AF271235 Rattus norvegicus dif
gb_pat:AB032435	+ 963.50	1499.10	3.9e-75	3346	AB032435 Homo sapiens DNPI MRN
gb_in:AV069501	+ 946.50	1475.87	7.6e-74	2748	AY069501 Drosophila melanogast
gb_ro:RN007609	+ 933.50	1458.41	7.2e-73	2024	U07609 Rattus norvegicus brain
gb_pr:AB032436	+ 932.00	1454.42	1.2e-72	2366	AB032436 Homo sapiens BNPI MRN
gb_pat:140029	+ 931.00	1451.41	1.8e-72	2716	I40028 Sequence 1 from patent
gb_pat:140028	+ 931.00	1451.41	1.8e-72	2716	I40029 Sequence 3 from patent
gb_pat:140220	+ 931.00	1451.41	1.8e-72	2716	I40220 Sequence 1 from patent
gb_pat:173259	+ 931.00	1451.41	1.8e-72	2716	I73259 Sequence 3 from patent
gb_pat:173260	+ 931.00	1451.41	1.8e-72	2716	I73260 Sequence 3 from patent
gb_in:AF024691	+ 907.50	1401.04	1.1e-69	9703	AF024691 Drosophila ananassa
gb_hlg:AC020077	+ 903.00	1377.50	2.3e-68	47223	AC020077 Drosophila melanogast
gb_in:AC091501	+ 903.00	1362.83	1.5e-67	198390	AC091501 Drosophila melanoga
gb_in:AC003806	+ 903.00	1359.72	2.3e-67	268219	AE003806 Drosophila melanoga
gb_in:CEC38C10	+ 892.00	1363.49	1.4e-67	34193	Z19153 Caenorhabditis elegans
gb_ro:BC018306	+ 885.50	1383.09	1.1e-68	1874	BC018306 Mus musculus, Similar
gb_pr:AB020527	+ 874.50	1362.18	1.6e-67	2626	AB020527 Homo sapiens mRNA for
gb_pr:AK024503	+ 864.50	1342.02	1.9e-66	3616	AK024503 Homo sapiens cDNA: FI
gb_in:AY075277	+ 859.50	1343.81	2.0e-66	1710	AY075277 Drosophila melanogast
gb_hlg:AC014600	+ 859.50	1311.95	1.0e-64	34229	AC014600 Drosophila melanogast
gb_in:AE003640	+ 859.50	1290.77	1.6e-63	267488	AE003640 Drosophila melanoga
gb_in:DROSADH01	+ 859.50	1289.46	1.8e-63	303885	AE003407 Drosophila melanoga
gb_in:AC023715	+ 845.50	1273.83	1.4e-62	160622	AC023715 Drosophila melanoga
gb_in:AE003432	+ 845.50	1266.74	3.4e-62	319551	AE003432 Drosophila melanoga
gb_hlg:AC013869	+ 829.50	1275.55	1.1e-62	11568	AC013869 Drosophila melanogast
gb_hlg:AC020252	+ 827.00	1259.66	8.4e-62	36842	AC020252 Drosophila melanogast

gb_in:AC007756 - 827.00 1243.90 6.4e-61 170105 ! AC007756 Drosophila melan
gb_in:AC007757 - 827.00 1243.85 6.4e-61 170939 ! AC007757 Drosophila melan
gb_in:AE003730 - 827.00 1241.19 9.0e-61 221409 ! AE003730 Drosophila melan
gb_pr:AK025880 + 826.00 1290.46 1.6e-63 1587 ! AK025880 Homo sapiens CDNA:
gb_in:AF095987 + 819.00 1276.10 1.0e-62 2178 ! AF095987 Caenorhabditis ele
seq_name: gb_om:AF244578
seq_documentation_block:
LOCUS AF244578 2844 bp mRNA linear MAM 06-AUG-2000
DEFINITION Ovis aries membrane glycoprotein SP55 (sp55) mRNA, complete cds.
ACCESSION AF244578
VERSION AF244578.1 GI:9719375
KEYWORDS sheep.
SOURCE
ORGANISM
Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Ovis.
1 (bases 1 to 2844)
Fu.C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P.,
Carter,C.E., Shi,E., Venkov,C., Yakes,M.F., Page,D.L. and H.C.G.
Identification of a novel membrane protein from mammalian cells
that interact with the anti-pathogenic compound CM101
Unpublished
2 (bases 1 to 2844)
Fu.C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R. and Heltqvist,C.G.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Location/Qualifiers
1..2844
/organism="Ovis aries"
/db_xref="taxon:9940"
/cell_type="endothelial cell"
/tissue_type="lung"
1..2844
/gene="sp55"
84..1571
/gene="sp55"
/note="similar to Na+ inorganic phosphate transporter"
/codon_start=1
/product="membrane glycoprotein SP55"
/protein_id="AAF9770.1"
/db_xref="GI:9719376"
/translation="MKSPVSLAPSDGEGSDRPTLQRAEAPVCCSARYNLAF
LSFGFFVLYSLRVNLSVALVDMSNTAKDRTSYEAHSAPFKVHQTGKKYR
WDAETQGLIGSFYGYITIQPGYVASRGGKLLGFGIFATAIFLFTPLAADFG
VGLVALRALGELGEGVITYPAMHMSWAPLERSKLLSISYAGAQLTGYVSLPSG
VICYMMNTYVYFPGIWIWILICVSDPTHTKTIPTPEKEYILSLKNQLS
QKSPWIPMLKSLPLWATVAHFYNWTFYLLTLLPTMYKVLRFQENGFLSAPV
YLGWLCLMSQADNLRAWNEFWLWVRVSLIGMIGPAFLVAAGFIGCDYSLA
VAFITISITLGGFCSSGFSINHLDIAPSYAGILLIGITWTFATIGMIGFIARSITPE
NTIGEQVTFICIAAINVFAIGFTLFKAGEVQNWAISSDHGHRN"

BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN
alignment_scores:
Quality: 2617.00 Length: 495
Ratio: 5.287 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-776-865-4 x AF244578
Align seg 1/1 to: AF244578 from: 1 to: 2844
1 MetLysSerProValSerPLeuAlaProSerAspGlyGluGlySe 17
|||||
84 ATGAAGTCCCGGTTTCGGACTTAGCCCCGACGCGGAGGAGGCTC 133
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProv 34

1034	GGAAGTCTCAAGTTCAATATTCAAGAAATGGGTTTTTATCTGCAGTCC	1089
334	rOfyrlLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAAsp	350
1084	CtTAUTTAGGTGTGGTTATGTATGATCATCTGTGGGTCAAGCTGCTGAC	1133
351	AnnLeuArgAlaArqTrpAsnPheSerThrLeuTrpValArgArgValPh	367
1134	AATTTAAGGGCAAGATGGAATTTTTCAACTCTGTGGGTTCGAAGAGTTTT	1183
367	eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP	384
1184	TAGCCTTATAGGGATGATTGGACCTGCATATTCTGGTGGCCGAGAT	1233
384	heIlelCyAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr	400
1234	TTATAGGCTGTCAATTATTCCTGGCTGTGCATTCCTAACCATATCAACA	1283
401	ThrLeuGlyGlyPheCySerSerGlyPheSerIleAsnHisLeuAspIl	417
1284	ACCCTGGGAGGCTTTTGCTCTCTGTGGATTATGACATCAACCATCTGGACAT	1333
417	eAlaProSerTyralaGlyIleLeuLeuGlyIleThrAsnThrPheAlar	434
1334	TGCTCTCTGATAGCTGGTATTCTCCTGGGCGATCACAAATACCTTTGCCA	1383
434	hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu	450
1384	CTATCTCTGGAATGATTGGGCCCATCATTCGCCAGAAGCTTACCCCTGAG	1433
451	AnnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs	467
1434	AACACTATTGGAGAATGGCAAACCTGTTTCTTCGATCCTGCTGCTATCAA	1483
467	nValPheGlyAlaIlePheThrLeuPheAlaLysGlyGluValGlnA	484
1484	TGTATTGGTGCCATTTCTTCACACTATTGCCCAAGGTGAAGTGCNAA	1533
484	sNrPalalaleserAspHisGlnGlyHisArgAsn	495
1534	ACTGGGCCATCAGTGCATGCCAAGGACACAGAATC	1568

seq_documentation_block:	
LOCUS	AX207626
DEFINITION	Sequence 3 from Patent WO0156598.
	2844 bp
	DNA
	linear
	PAT 31-AUG-2001

ACCESSION	AX207626	
VERSION	AX207626.1	GI:15422331
KEYWORDS		
SOURCE	Ovis sp.	
ORGANISM	Ovis sp.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae; Bovidae; Caprinae; Ovis.	
AUTHORS	1 (bases 1 to 2844)	
TITLE	Hellerqvist.C.G.	
JOURNAL	Methods for preventing or attenuating pathoangiogenic conditions by using the gbs-toxin (cm101) receptor as a vaccine	
	Patent: WO 0156598-A 3 09-AUG-2001.	

```

FEATURES             source          CDS
VANDERBILT UNIVERSITY (US)
Location/Qualifiers
1..2844
/organism="Ovis sp."
/db_xref="taxon:9939"
84..1571
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60359.1"
/db_xref="GI:15422332"
/translators="MKSPVSDIAPSDEGEISDRTPLLQRPRAEPAPVCVCSARYNLAF
LSPFETGVLYSLRVLSVALVDVNDVNTTAKDNRTSYECAHESAPIKVLHNQTKKYYR
WDFGFGWGLSGFEYGVITTOIPGGVVASRSGKLLIGCFIPATFTFTLTPDAADFGR

```

VGALVALRALEGLGEGVTYPAMHAMSSWAPPLERSKLLSISYAGAQLGTVVSLPLSG
VICYMNNTYVFFYFGIIVFIWILMCLIVSTPETHKTIIPYERKEYILSSKLNOLSS
OKSVPIWMLKSLPLWAIWVAFHSYNWTFYLLTLPTVMKVLRFNIOENGFLSAVP
YLGCVLWCLNLSGOADNLRARNFSTLWVRVFSLIGMIGPAIFLVAAGFICDYSLA
VAFLLISITLGGFCSSGFSINHLIDAPSFAGILLGTFNFATIPGMIGFPIIARSLTPE
NTFISGMQVFCIAAAINVFAGITFFFLFARKEVQONWAISSHQHRN"

BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN

alignment_scores:
Quality: 2617.00 Length: 495
Ratio: 5.287 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-776-865-4 x AX207626 ..

Align seg 1/1 to: AX207626 from: 1 to: 2844

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||
84 ATGAAGTCCCGGTTTCGGACTAGCCCCGCGAGCGCGGAGGAGGCTC 133
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
|||||
134 GGACCGCACACCGCTCTCGAGCGCGCGCGCGGCGGAACCCGCTCCAG 183
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
|||||
184 TATGCTGCTGCTGCTGCTTAAACCTAGCATTTTGTCTCTTTTGGTTTC 233
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
234 TTCGTTCTCTATTCATTACGGGTGAATCTGAGCGTTGCACCTAGTGGACAT 283
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
284 GGTGGATTCAACACAACTGCCAAAGATAATAGAACCTCCTACGAGTGTG 333
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
334 CAGAGCATCTGCTGCCATAAAAGTTCTTCAACACCAACCGGGTAAAAAG 383
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
384 TACCGGTGGGATGCAGAAACTCAAGATGGATTCCTGGATCTTTTCTA 433
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerg 134
|||||
434 TGGCTACATCATCACAAAATCTCTGGAGGATATGTTGCCAGCAGAAAGTG 483
134 lyGlyLysLeuLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
484 GGGGGAAGCTGTGCTAGGATTCGGGATCTTGCACAGCTATCTTCACC 533
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
534 CTGTTCACTCCCTCGCTGCAGATTCGGAGTCCGAGCCCTTGTGGCACT 583
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethisA 184
|||||
584 CAGGGCACTAGAAGGGCTAGGAGGGGTGCACATATCCAGCCATGCATG 633
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
634 CCATGTGGTCTTATGGGCTCCCTCTTGAAGAAGCAAGCTCTCTGAGT 683
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
684 ATTTTCATATCAGGAGCACAACTTGGACAGATAGTTCTCTCTCTCTTC 733
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234

|||||
734 TGGAGTAATTTGCTACTATATGAATTGCATTTATGTCTTCTTATTTCTTTG 783
234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
784 GCATTTGTTGAATCATCTGGTTTATTTTATGGATCTGCTTAGTTAGTAT 833
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
834 ACACCAGAAACTCACAGACAATCACTCCCTATGAAAGGAGTATATTTCT 883
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
884 TTCATCATTAATAAATCAGCTCTCTTACAGAACTCAGTGCCTGGATAC 933
284 roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
|||||
934 CTATGCTGAAATCACTGCCACTTTGGGCTATTGCTGTTGCACATTTTCT 983
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||
984 TACAACCTGGACTTTTATACTTTGTCACCTTATGCTACTTACATGAA 1033
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
1034 GGAAGTCTTAAGGTTCAATATTTCAAGAGAAATGGGTTTTTATCTGCA 1083
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||
1084 CTTATTTAGGTTGTTGTTATGATGATCTGCTGGGTCAAGCTGGTGAC 1133
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1134 AATTTAAGGCAAGATGGAATTTTCAACTCTGTGGGTTTGAAGAGTTTT 1183
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1184 TAGCCTTATAGGATGATTTGGACCTGGGATATTCCTGTTGCCGAGGAT 1233
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1234 TTATAGGCTGTGATTTATCTCTTGGCTGTGCATTTCTTAACCATATCA 1283
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1284 ACCCTGGAGGCTTTTCTCTCTGGATTTAGCATCAACCATCTGGACAT 1333
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAla 434
|||||
1334 TGCTCTTCGTATGCTGGTATTCCTTGGGCATCACAATACCTTTGCCA 1383
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1384 CTATTTCTGGAATGATTTGGGCCCATCATTTGCCAGAGCTTTACCCCTG 1433
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaAlaIleAs 467
|||||
1434 AACACTATTGGAGAAATGGCAAACTGTTTCTGTCATCGCTGCTATCAA 1483
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGln 484
|||||
1484 TGTATTTGGTGCCATTTTCTTCACTATTTCCGCAAGGTGAAGTGCAAA 1533
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||
1534 ACTGGGCCATCAGTGTATCACAAGGACACAGAAAC 1568

seq_name: gb_pat:AX138494

seq_documentation_block:

LOCUS AX138494

DEFINITION Sequence 2 from Patent EP1069184.

ACCESSION AX138494

linear

PAT 30-MAY-2001

```
VERSION      AX138494.1  GI:14274389
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 2512)
AUTHORS      Human anion transporter gene implicated in salla disease and
              lysosomal sialic acid transport
TITLE        Patent: EP 1069184-A 2 17-JAN-2001;
              Akzo Nobel N.V. (NL)
JOURNAL
FEATURES
source       Location/Qualifiers
              1..2512
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   612 a 585 c 566 g 749 t
ORIGIN
alignment_scores
Quality: 2329.00      Length: 495
Ratio: 4.903          Gaps: 0
Percent Similarity: 95.960  Percent Identity: 86.263
alignment_block
US-09-776-865-4 x AX138494 ..
Align seg 1/1 to: AX138494 from: 1 to: 2512
1  MethySerProValSerAspLeuAlaProSerAspGlyGluGluGlyse 17
|||||
271  ATGAGGCTCCGGTTCGAGACGTCGCCGGAACGATGCGGAGGAGCAC 320
17  rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
|||||
321  GGACCGCACGCGCTCTTACCGCGGCGCCACGGCGGCGGACCGCTCCAG 370
34  aLysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
|||||
371  TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
51  PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
421  TTCATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 470
67  tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
471  GGTAGATTCAAATCAAACTTTAGAGGATTAATAGAACTTCCAAGCGGTGC 520
84  LaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
521  CAGAGCATTCGCTCCCATAAAGTTTCATCATCAATCAACCGGTAAGAG 570
101  TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
571  TACCAATGGGATGAGAACTCAAGAGTGGATTCTCGGTTCTCTTTTAA 620
117  rGlyTyrIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
621  TGGCTACATCATCATCAGATCTCTGGAGGATATGTTGCCAGCAAAATAG 670
134  lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
671  GGGGAAATCTGCTAGGATTTGGGATCTTGGCACTGCTGCTCACC 720
151  LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
721  CTGTTTCATCCCATTTGTCAGATTTAGAGGTTGGACCTCACTGTACT 770
167  uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
|||||
771  CAGAGCATTAGAGGATAGGAGGAGGTTTACATTTCCAGCCCATGCATG 820
```

```
184  laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
821  CCATGTGTCTCTCTTGGCTCCCTCTTGAAGAAGCAAACTTCTTAGC 870
201  IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
871  ATTTTCATATGAGGAGCAGACAGCTTGGGACAGTAATTTCTCTCTCTTC 920
217  rGlyValIleCysTyrTrpMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||
921  TGGAAATAATTTGCTACTATATGAATGGACTTGTCTCTTACTTTTTTG 970
234  lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
971  GTACTATTTGGAAATATTTGGTTCTTTTGGGATCTGGTTAGTTAGTGAC 1020
251  ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
1021  ACACCACAAAACACAAGAGAATTTCCCATTTATGAAAAGGAATACATTCT 1070
267  uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
1071  TTTCATCATTAAGAAATCAGCTTTCTTACAGAACTCAGTCCCGTGGGTAC 1120
284  roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
|||||
1121  CCATTTTAAATCCCTGCCACTTTGGGCTATCGTAGTTGCACACTTTTCT 1170
301  TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||
1171  TACAACCTGGACTTTTATCTTTATGACATTTATGCTTACTTATATGAA 1220
317  sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
1221  GGAGATCCTAAGGTTCAATGTTCAGAGAAATGGGTTTTTATCTTCATGTC 1270
334  roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||
1271  CTATATTAGGCTCTGTGTTATGATGATCCTGCTGCTGCTGCTGCTGCTG 1320
351  AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1321  AATTAAAGGGCAAAATGGAATTTTCAACTTTATGTTGCTGCGAGAATTTT 1370
367  eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1371  TAGCCTTATAGGAATGATGTACCTGCAGATTCCTGCTGCTGCTGCTGCT 1420
384  heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1421  TCATTGGCTGTGATTATTTTGGCCGCTGCTTTCTTAACTATATCAACA 1470
401  ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1471  ACATGGGAGGCTTTTGTCTCTCTGAGTATTTAGCATCAACCATCTGGAT 1520
417  eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||
1521  TGCCTCTGTATGCTGATCCTCCCTCCGCGCATCACAAATACATTGCCA 1570
434  hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1571  CTATTCCAGGAATGGTTGGCCCGCTCATTTGCTAAAGCTGACCCCTGAT 1620
451  AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1621  AACACTGTTGGAGAAATGGCAACCGCTGTTATATATGCTGCTATATAA 1670
467  nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1671  TGTTTTGGTGCCATTTCTTTACACATTTCCGCCAAAGGTGAAGTACAAA 1720
```


seq_documentation_block:
LOCUS AF244577 2930 bp mRNA linear PRI 06-AUG-2000
DEFINITION Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
ACCESSION AF244577
VERSION AF244577.1 GI:9719373
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2930)

AUTHORS FU,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P., Carter,C.E., Shi,E., Venkov,C., Yakes,M.F., Page,D.L. and Hellerqvist,C.G.
TITLE Identification of a novel membrane protein from mammalian cells that interacts with the anti-pathangiogenic compound CM101 unpublished
JOURNAL 2 (bases 1 to 2930)
REFERENCE FU,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R. and Hellerqvist,C.G. Direct Submission
TITLE Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School of Medicine, 23rd pierce, Nashville, TN 37232-0146, USA
JOURNAL of Medicine, 23rd pierce, Nashville, TN 37232-0146, USA
FEATURES
 source 1. .2930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 gene 1. .2930
 /gene="HP59"
 CDS 263..1873
 /gene="HP59"
 /notes="similar to Na+ inorganic phosphate transporter"
 /codon_start=1
 /product="membrane glycoprotein HP59"
 /protein_id="AAF97769.1"
 /db_xref="GI:9719374"
 /translation="MAAGATPPRPQVPARGPGGLSGRRSLQVASTPAHVGVNMRSPVLDLRNDEESTDRTLLPGAPRAEAPVCNSARYNLAILAFFGFFIVYALRVNLISVALVMDVNSTLTEDNRTSKACPEHSAPIKVHNHNTGKYOWDAETQGTWLTGSGFFGYILITQIPGGYVASKIGKMLGFCILGTAVLTFTPTAADLGVGLIVLRALEGIGGVTPFAMHMSWAPPLERSKLISYAGAQLGTVISLPLSGIICYANNTWTVYVYFPGTIGIFWELLTWLSDTPQKHRSHEYKEYILSLRNQLSSQSKSPWVPVILKSLPLWLAIVVAHFSYNNWTFLLTLLPYMKEILRFNVQENGLSSPYLGSLWCLMILSQAAADLRKAWNESTLCVRIFSLIGMIGPVLVAAGTIGDYSLAVAFLLTSTTLGGFCSSGFSINHLIDAPSYAGILGINTPATIPGMVGPVIAKSLTDPDNTVGEMQTVFYIAAAILVFGAIFPTLFAKGEVQWALNDHGHRRH"
BASE COUNT 729 a 698 c 681 g 822 t
ORIGIN

 alignment_scores:
 Quality: 2329.00 Length: 495
 Ratio: 4.903 Gaps: 0
 Percent Similarity: 95.960 Percent Identity: 86.263

 alignment_block:
 US-09-776-865-4 x AF244577 ..

 Align seg 1/1 to: AF244577 from: 1 to: 2930

 1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlyse 17
 ||||:||||| ||||| :|||:|||||:|||||
 386 ATGAGGTCTCCGGTTCGAGACCTGCGCCGGAACGATGCGGAGGAGACAC 435

 17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
 :||||| ||||| ||||| |||||
 436 GGACCGCACCGCTCTCTACCGGGCGCCGCCAGCGGCGCAAGCGCGTCCAG 485

 34 aLCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
 :||||| ||||| ||||| |||||
 486 TGTGCTGCTCTGCTCTGTACAACTTAGCAATTTTGGCCTTTTGGTTC 535

 51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
 ||||:||||| ||||| ||||| |||||
 536 TTCATGTGTATGCATACGTGTGAATCTGAGTGTGCGTGTAGTGGATAT 585

 67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
 ||||:||||| ||||| ||||| |||||
 586 GGTAGATTCAATAACAACCTTAGAAGATAATAGAACCTTCCAAGCGGTGTC 635

 84 IaGluHisSerAlaProIleLysValLeuHisAsnGlnThrClyLysLys 100
 ||||:||||| ||||| ||||| |||||
 636 CAGAGCATCTGCTCCCAAAAAGTTTCATCATTAANTCAACACGGGTGAAG 685

 101 TVrArTrpAspAlaGluThrGlnGlyTrpIleLeuGlyVserPhePhePhe 117

/co
/pr

CDS


```

transporter), member 5"
/protein_id="AAH20961.1"
/db_xref="GI:18088703"
/translation="MRSVPNDLARNDGEESTDRTPLLPGAPRAEAAAPVCCSARYNLAI
LAFEGFIVYALVNSILVALVDMSVNTILEDNTKACPEHSAPTKVHHNQTKKKYQ
WDATOGWILGSFFYGTITIPGGVAKIGKMLGPGILGTAVLTLPPTAADLG
VGPLIVRLALEGVEFFFGTIPFAMHAWSSWAPPLERKSLISISYAGQLGVISLPJSS
IICYVNLNIVYFFFGTIGTFWFLMIWLVDVTPQKHKRISHYEYKIVLSSLRQSLG
QKSPVQMLAANLVAHVSYNWTFITLLTPYKMEILRNQVQNGFSLSP
YLGSLWCMILSGOANDLNAKWNFSTLCVPRFSLIGMTGIPVFLVAAGPIGCDYSLA
VAPLTISTTLGGRCSSGFSINHLDIAPSLAGILLGITNTFATTPGMPGVPIAKSLTPD
NTVGEWQTVFYIAAANVFGAIFFTLFCAGVEVQNALNDHGHRR"
863 a 718 c 727 g 1021 t

```

[illegible]

seq_name: gb_in:AY060776

seq_documentation_block:

LOCUS AY060776 1841 bp mRNA linear INV 08-NOV-2001

DEFINITION Drosophila melanogaster GH23975 full length cDNA.

ACCESSION AY060776

VERSION AY060776.1 GI:16768209

KEYWORDS FLI_CDNA.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1841)

Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,

Champe,M., Chavez,C., Dorsett,V., Fafán,D., Frise,E., George,R.,

Gonzalez,M., Guarín,H., Li,P., Liao,G., Miranda,A., Mungaii,C.J.,

Nunoo,J., Pacile,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,

Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.

Direct Submission

Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,

Lawrence Berkeley National Laboratory, One Cyclotron Road,

Berkeley, CA 94720, USA

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to

sequence clones from Drosophila Gene Collection 1 (Rubin et al.,

Science 2000). The sequence has been subjected to integrity checks

for sequence accuracy, presence of a polyA tail and contiguity

within 100 kb in the genome. Thus we believe the sequence to

reflect accurately this particular cDNA clone. However, there are

artifacts associated with the generation of cDNA clones that may

have not been detected in our initial analyses such as internal

priming, priming from contaminating genomic DNA, retained introns

due to reverse transcription of unspliced precursor RNAs, and

reverse transcriptase errors that result in single base changes.

For further information about this sequence, including its location

and relationship to other sequences, please visit our web site

(<http://fruitfly.berkeley.edu>) or send email to

cdna@fruitfly.berkeley.edu.

Location/Qualifiers

1..1841

/organism="Drosophila melanogaster"

/strain="v: cn bw sp"

/db_xref="taxon:7227"

/map="92E8-92E8"

1..1841

/gene="CG4288"

/note="alignment with genomic scaffold AE003730"

/db_xref="FLYBASE:FBgn0038799"

133..1641

/gene="CG4288"

/note="Longest ORF"

/codon_start=1

/db_xref="FLYBASE:FBgn0038799"

/product="GH23975"

/protein_id="AAL28324.1"

/db_xref="GI:16768210"

/translation="MTDLESPKNGKHLEQVHNNNEITSEPLTLRFRWKORYIVVLL

APFGFNYSVRLNLSVAIVMTENRTVFDADGNVSYQDDPWSKQKGLILSSPYG

YLTQFLGGYIGTKIGNVFTGIGSTAILTLTPMAASHLEMLFVRLIEGFEG

VTFPGIHAVARWSPPLERSRWASIAFAGNAGTVVAMPSCSFLATKYGWESVYVFG

TIGVLIWTLVYVFKAGPELDRFCSEEDYIKQTIQYVGSKHVHPWRAIFTSMPFY

AIMASSENNGYFLLTLQPLSFRLDNLFDLQKTIILSAVPYLAAGLLAVSGYIAD

WQVQGIWTTQVRRNFNCFAFIAQTFWMLTAYLLDPTWSVSLTIAGVGAFAWSG

FAVNHLDIAPQHASVLMGIGNTFATIPGVISPLTGYVVTNOTSEWRIIFISAGIY

LVGCVIYWFYCSGDLOEWAKTPEQAKAEKAEKALQITQTAGFVNSGAEIKD"

392 a 495 c 479 g 475 t

BASE COUNT
ORIGIN

alignment_scores:

Quality: 985.50 Length: 499

Ratio: 2.730 Gaps: 6

Percent Similarity: 72.345 Percent Identity: 39.078

alignment_block:

US-09-776-865-4 x AY060776 ..

Align seg 1/1 to: AY060776 from: 1 to: 1841

5 ValSerAspLeu...AlaProSerAspGlyGlu..... 14

133 ATGACCCAGCCTGGAGTCTCCGAGAACGGAACACCTGGAGCAGGTGCA 182

15GluGlySerAspArgThrProLeuLeuGlnArgAlap 27

183 CCAAAACAACATGAATACGGAGAGTGACCACTCACCTGGCGGATTTT 232

27 roArgAlaGluProAlaProValCysSerAlaArgTyrAsnLeuAla 43

233 GCGCTAAACAG.....CGCTACATTGTGGTG 258

44 PheLeuSerPhePheGlyPheValLeuTyrSerLeuArgValAsnLe 60

259 CTACTGGCTTTCTTCGGATTCTTCAATGCTACTCACTGGCGGTAATCT 308

60 uSerValAlaLeuValAspMetValAspSerAsnThrThr...AlaLys 76

309 TTCGGTGGCATTGTGCCATGACCGAGAAATCGCACCGCTTCGTGATCGCG 358

76 sPasnArgThrSerTyrGluCysAlaGluHisSerAlaProIleLysVal 92

359 ATGGAATGTGCTCTACAG..... 378

93 LeuHisAsnGlnThrGlyLysLysTyrArgTrpAspAlaGluThrGlnGl 109

379CAGGACTTTCCCTGGGAGCTCCCAAGCAGAGGG 410

109 yTrpIleLeuGlySerPhePheThrGlyTyrIleIleThrGlnIleProG 126

411 CTGTGATCCTGAGCTCTTCTTACGGCTACATCCTGCACAGTTCCTCG 460

126 lyGlyTyrValAlaLaserArgSerGlyGlyLysLeuLeuLeuGlyPheGly 142

461 GTGGCTACATCGGCACCAAAATGGTGCACACATTGATTGGTACTGGA 510

143 IlePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPh 159

511 ATTGGAAGACACTGCCATTCTGACGCTGCTCACTCCGATGGCTCGCGCA 560

159 eGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluG 176

561 CAGCTTGGAAATGTCTCTGTCTGTCGCGCATCATCGAGGGCTCTCTCGAGG 610

176 lyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaProPro 192

611 GCGTCACTTCCCGGTATTCATCGCGCTCTGGCCCGCTGGTCTCCGCT 660

193 LeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGl 209

661 TTGGAACGATCCCGAATGGATGCATAGCATTTTCCCGCAACTATGGGG 710

209 yThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMetAsnT 226

711 AACGGTGGTGGCGATGCATGCTCCGGATTCTGGCTTACCAATACGGCT 760

226 rpThrTyrValPheThrPheGlyIleValGlyIleIleIleTrpPheIle 242

761 GGAGACGCGTATTCTACGTATTGGCACCATCGCGGTGATCTGGTACATC 810

243 LeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrIleTh 259

811 ACTTGGCTCGTCTCGTCAAAAGCTGGACCGAGCTGGATCGCTTCTGTTTC 860

6386 ATCCAGCTATGACACCCATGCTGCCACTGGATACCCCGCTGGAGAG 6435
196 SerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValVa 212
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6436 AACAGTTGCGGCAATCGTATGCGGGCTCCAAATATCGGAACAGTCAT 6485
212 IserLeuProLeuSerGlyValIleCys.....TyrTyrMetAsnT 226
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6486 TTCATGCGCTGCGCGATGCTGTGCTGCTGGACTTCCTGGGTGGCT 6535
226 rpThrTyrValPheTyrPheGlyIleValGlyIleIleTrpPheIle 242
||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6536 GGCGCTCGGCTTCTACATCTTCGAGCTGTGGCATCTCTGTGTTTCATC 6585
243 LeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrIleTh 259
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6586 GCATGGATGTTATTTGGTGTACGAAGCTAGCGATCATCCAGGATCTC 6635
259 rProTyrGluLysGluTrpIleLeuSerSerLeuLysAsnGln..... 273
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6636 AGAATCGGAGCGAGATATATCGAAGAGTCTACAGGTTCCAGAGCTAA 6685
273 273
6686 TAAATCAGGATCTAGCGGAGCGGAGGAGGAGGAGGAGGATGAAGTG 6735
274LeuSerSerGlnLysSerValProTrpIleProMetLeuLy 287
6736 AGTTCGGGGCGCCCGGAGAACCGATACCTGGTGCATCGTCTGCAC 6785
287 sSerLeuProLeuTrpAlaIleValAlaHisPheSerTyrAsnTrpT 304
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6786 ATCCGTACCTCTGTGGGCCATCTTTGTCACGCAATGCGGCCGAGGATGG 6835
304 hrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGluValLeu 320
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6836 CTTTACACGACGATACCGAGTGTCCACCTACATGAGCAACATCCTA 6885
321 ArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLeuG 337
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6886 CACTTGCATCCAGTCGAATGCTGTGCTCAATGCGGTGCGGTATCTAAC 6935
337 yCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArg 354
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6936 CTCCTGGTTCTGGCATGCTGCTGCTCGCCCTGCGGATGGATGCTAG 6985
354 laArgTrpAsnPheSerThrLeuTrpValArgArgValPheSerLeuIle 370
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6986 CCAGACGCTACATATCGCTGCTGAACCTCGTATAAGTTGTGGAACCGGTG 7035
371 GlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGlyCy 387
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7036 CCTTCGGTGGTCCATCACTGGGCTGATGGCATCATCTATGTGGGCTG 7085
387 sAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyG 404
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7086 CGATTGGGTGGTGCACCTTATGTTGGCGCGGTGGCTCTTTCGGCG 7135
404 lypPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSer 420
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7136 GCGCGCTCTATGCTGGCAACAGATGAATCATATAGCGCTCAGTCCACGA 7185
421 TyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProG 437
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7186 TATCAGGACCATGATGGCATCATCACCATTTCGGCGCAAAATATCTGTGG 7235
437 yMetIleGlyPro...IleIleAlaArgSerLeuThrProGluAsnThrI 453
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7236 CTTCTGCTGCTGATGTCATCGTCTTAATCATCAATCATCGCGAGACTC 7285
453 leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7286 TGACACAGTGGCATCTGCTCTTCTGGTGGCGGCGGCTTGAATATAGCC 7335

470 GlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAsnTrpAl 486
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7336 GGTAACCTTCATCTACCTATCTTCCAGCCGCCGAGGAGCAAGCTGGTC 7385

486 a 486
7386 G 7386

seq_name: gb_in.AC023685

seq_documentation_block:

LOCUS AC023685 167928 bp DNA linear INV 03-JAN-2002
DEFINITION Drosophila melanogaster 3L BAC RP98-20N12 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.

ACCESSION AC023685
VERSION AC023685.3 GI:17223100
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 167928)

AUTHORS

Gocayne,J.D., Tabor,P., Williamson,A., Homsí,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayale,M.A., Scott,G.S., Worley,K.W., Amaratunga,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Fertiera,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattel,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 167928)

Worley,K.C.

Direct Submission

Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Paylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 167928)

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozago,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

4 (bases 1 to 167928)
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

5 (bases 1 to 167928)
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE

JOURNAL

Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

6 (bases 1 to 167928)
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

alignment_scores:

Quality:	981.50	Length:	517
Ratio:	2.853	Gaps:	6
Percent Similarity:	66.538	Percent Identity:	38.685

alignment_block:

```
US-09-776-865-4 x AC023685  ..
Align seg 1/1 to: AC023685 from: 1 to: 167928
3 SerProValSerAspLeuAlaProSerAspGlyGluGlySerAspAr 19
||| :|||:||||| |||:|||||: |||: |||
149662 TCCACTCTCTCTCTCCACCCTGCAGAAACCATGCTGTGGCCCAA 149711
19 gThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProValCysC 36
:||||
149712 GACG..... 149715
36 ysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhePheVal 52
|||: :|||:|||||:|||||:|||||: |||
149716 .....CGTCACATTTTCGGATTTCATGGCTTCTTGGGATTCGCCGTG 149757
53 LeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMetValAs 69
:||||:|||||:|||||:|||||:|||||: |||
149758 GTCTACGGCATGCGGGTCAATCTGTGGTGGCCATTGTGGCCATGGTG... 149805
69 pSerAsnThrThrAlaLysAspAsnArgThrSerTyrCluCysAlaGluH 86
||| |||||: |||: |||: |||
149806 ....AACCAACGGCAATCCGCACACCACTCATCGGTGATTGATACGG 149851
86 isSerAlaProIleLysValLeuHisAsn..... 95
:||||:|||||: |||: |||: |||
149852 ACAGTGTCCTACTACCGGCACACACATCACAATGGTAGGATCCCAATCCG 149901
96 GlnThrGlyLysLysTyrArgTyrAspAlaGluThrGlnGlyTrpIleLe 112
||| :|||: |||: |||: |||: |||
149902 CAGAAGGAGGCGAGTTGTGTGGAGGAGGCCACGCGGATTTGGTGTCT 149951
112 uGlySerPhePheThrGlyTyrIleIleThrGlnIleProGlyGlyTyr 129
|||||:|||||:|||||:|||||:|||||: |||
149952 CGCGAGTTTCTTCATGCTATGTGTGCTAAACCAAGTCCCGCGGCGGAGA 150001
129 aAlaSerArgSerGlyLysLeuLeuGlyPheGlyIlePheAla 145
:||||: |||: |||: |||: |||: |||
150002 TGGCGGAGCTGTATGTTGGGAAGAAGATCTACGGCTATGGAGTCTGTATC 150051
146 ThrAlaIlePheThrLeuPheThrProLeuAlaAlaaspPheGlyValGI 162
|||||:|||||:|||||:|||||:|||||: |||
150052 ACGGCGTCTTTACGCTTATAACCTTCAATGGCTGCCACTGGGATCTGCC 150101
162 yAlaLeuValAlaLeuArgAlaLeuGlyGlyLeuGlyGlyValThrT 179
||||| :|||: |||: |||: |||: |||
150102 GCTGTTGGTCTGTGTCGCATCTTGGAGGGAATGGCGAGGCGGTCACT 150151
179 yrProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGluArg 195
|||||:|||||:|||||:|||||:|||||: |||
150152 ATCCAGCTATGCACGCCATGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGC 150201
196 SerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValVa 212
:||||: |||: |||: |||: |||: |||
150202 AACAAAGTTCGGCGCAATCGTCTATGCGGGGTCCCAATATATCGGAACAGTCAT 150251
212 lSerLeuProLeuSerGlyValIleCys.....TyrTyrMetAsnT 226
:||||:|||||:|||||: |||: |||: |||
150252 TTCCATGCGCGTCCCGGATGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 150301
226 rpThrTyrValPheTyrPheGlyIleValGlyIleIleTrpPheIle 242
||| :|||:|||||:|||||:|||||:|||||: |||
150302 GGCGGTGGCTTTCTACATCTTCGACTGTGTGGCATTCGTGTGTTTCATC 150351
243 LeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrIleTh 259
|||||:|||||: |||: |||: |||: |||
150352 GCATGGATGATTGTGTGTACGACAAAGCTAGCGATCATCCAGGATCTC 150401
259 rProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGln..... 273
:||||:|||||:|||||: |||: |||: |||
150402 AGAATCGGAGGAGTATATCGAAGAGGTCTACAGGTTTCAGAGGTAA 150451
273 ..... 273
```

```
150452 TAAATCAGGATCTAGCGGAGCCGAGGAGGAGGAGGACAGGATGAAGTG 150501
274 .....LeuSerSerGlnLysSerValProTrpIleProMetLeuLys 287
:||||: |||: |||: |||: |||: |||
150502 AGTCTGGGGCGCGCGGAGGAACCGATACCCCTGGTCTATCGCTGCTCAC 150551
287 sSerLeuProLeuTrpAlaIleValValAlaHisPheSerTyrAsnTrpT 304
|||||:|||||:|||||:|||||:|||||: |||
150552 ATCCGTACCTCTGTGGGCCATCTTGTGACGCAATCGCGCAGGATGGG 150601
304 hrPheTyrThrLeuLeuLeuLeuProThrTyrMetLysGluValLeu 320
|||||:|||||: |||: |||: |||: |||: |||
150602 CCTTCTACAGCAGCTAACCGAGCTGCCACCTACATGAGCAACATCCTA 150651
321 ArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLeuGI 337
:||||:|||||:|||||:|||||:|||||: |||
150652 CACTTTGACATCCAGTCGAATGCTCTCAATGCGGTGCGGTATCTAAC 150701
337 yCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgA 354
|||||: |||: |||: |||: |||: |||
150702 CTCTGTTTGTGGGCATTCCTCTCCGCCCTGGCGGATGGATGCTAG 150751
354 laArgTrpAsnPheSerThrLeuTrpValArgArgValPheSerLeuIle 370
||||| :|||: |||: |||: |||: |||: |||
150752 CCAGACGCTACATATCGCTGCTGAACCTGTAAGTTGTGGAACACCGGTG 150801
371 GlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGlyCy 387
: |||: |||: |||: |||: |||: |||
150802 GCCTCCGTGGTGGCATCCTGCGGCTGATTGGCATCATCTATGTGGCTG 150851
387 sAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyG 404
|||||: |||: |||: |||: |||: |||
150852 CGATTGGGTGGGTACCTTTATGTTGGCGCGGTGGCTCTTTCGGCG 150901
404 iYPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSer 420
||| :|||: |||: |||: |||: |||
150902 GCGCGCTATGCTGGCAACAGATGAATCACAATAGCGCTCAGTCCACGA 150951
421 TyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProGI 437
|||||: |||: |||: |||: |||: |||
150952 TATGACGACCATGATGATGCATCACAATTCGGCGGCAATATCTGTGG 151001
437 yMetIleGlyPro...IleIleAlaArgSerLeuThrProGluAsnThrI 453
:||||: |||: |||: |||: |||: |||
151002 CTTCCTGGCTCCGTATGCTATCGGTCTAATCATCAATCATCGCGAGACTC 151051
453 leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
: |||: |||: |||: |||: |||: |||
151052 TGACACAGTGGCATCTGCTCTTCTGCGGCGGCTTGAATATAGCC 151101
470 GlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAsnTrpAl 486
||| :||||: |||: |||: |||: |||: |||
151102 GGTAACCTTCATCTACCTGATCTTGGCCAGCGCGGAGGACAAAGCTGGTC 151151
486 a 486
151152 G 151152
```

seq_name: gb_in:AC023711

seq_documentation_block:

LOCUS AC023711 174163 bp DNA linear INV 19-DEC-2001
DEFINITION Drosophila melanogaster 3 BAC RP98-6C4 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.

ACCESSION AC023711

VERSION AC023711.3 GI:17933778

KEYWORDS HTG

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS

1 (bases 1 to 174163)
Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferreria,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Scheeler,F., Shen,H., Strong,N., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstein,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 174163)
Worley,K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

3 (bases 1 to 174163)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Blimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgatz,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.B., Jacobson,B., Jia,F., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Submitted (19-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL

On Dec 19, 2001 this sequence version replaced gi:6997282.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished,) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum

standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	Source	Location/Qualifiers
	1..174163	/organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="3" /clone="Rp98-6C4"
BASE COUNT	51926 a 36149 c 35565 g 50523 t	
ORIGIN		
alignment_scores:	Quality: 981.50 Length: 517 Ratio: 2.853 Gaps: 6 Percent Similarity: 66.538 Percent Identity: 38.685	
alignment_block:		
US-09-776-865-4 x AC023711		
Align seg 1/1 to: AC023711 from: 1 to: 174163		
3 SerProValSerAspLeuAlaProSerAspGlyGluGluGlySerAspAr 19		
: : : : : : :		
120244 TCCACTCTCTCTCTCCACCCTCGACAAACCATGGCTGGCCCAAA 120293		
19 gThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProValCysC 36		
: : : : : : :		
120294 GAGC.....		
36 ysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPheVal 52		
: : : : : : :		
120298CGTCACATYTCGGATTCATGGCTTCCTGGGATTCGCCGTG 120339		
53 LeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMetValas 69		
: : : : : : :		
120340 GTCTACGCGATCGGGTCAATCTCGGTGCCCATTTGGCCATGGTG.. 120387		
69 pSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysAlaGluH 86		
: : : : : : :		
120388AACCAACGGCAATTCGGCACACCACTCATCGGTGATTGATACGG 120433		


```

86 isSerAlaProIleLysValLeuHisAsn..... 95
   :::::||||:  ::::: |||:::
120434 ACACGTGTCCACTACCGCGCACACATACATGGTAGCATCCCAATCCG 120483

96 GlnThrGlyLysTyrArgTrpAspAlaGluThrGlnGlyTrpIleLe 112
   |||  ::::: |||||  |||||  ::::: |||
120484 CAGAAGGAGGGCGATTGTGTGGACGAGGCCACGACGAGATTGGTGT 120533

112 uGlySerPhePheTrpGlyTyrIleThrGlnIleProGlyGlyTyr 129
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
120534 CGCGAGTTCTTCTATAGCTATGTGCTTAACCAAGTCCCGCGGACGGA 120583

129 aLaSerArgSerGlyLysLeuLeuLeuGlyPheGlyIlePheAla 145
   ::::: |||:::  ::::: |||:::  ::::: |||:::
120584 TGGCCGAGCTGTATGGTGGGAAGAAGATCATCGGCTATGGAGTGTATC 120633

146 ThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPheGlyVal 162
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
120634 ACGCGGGTCTTTACGCTTATAACTCCATTTGGCTGCCCACTGGGATCTGCC 120683

162 yAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGlyValThr 179
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
120684 GCTGTGGTCTGTGCTGCCATCTCGAGGGAATGGCGAGCGGTCACT 120733

179 yProAlaMetHisAlaMetTrpSerTrpAlaProLeuGluArg 195
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
120734 ATCCAGCTATGCACGCCATGCTTGCCCACTGGATACCGCCGCTGAGAGG 120783

196 SerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrVal 212
   ::::: |||:::  ::::: |||:::  ::::: |||:::  ::::: |||:::
120784 AACAAAGTTTCGGCGCAATGCTATGCGGGCTCCCAATATCGGAACAGTCA 120833

212 LserLeuProLeuSerGlyValIleCys.....TyrTyrMetAsn 226
   ::::: |||||  ::::: |||  ::::: |||  ::::: |||
120834 TTCCATGCGCTGGCCGAGATGCGTGTGCTCGTGGACTTCTCTGGTGGCT 120883

226 rpThrTyrValPheTyrPheGlyIleValGlyIleIleTrpPheIle 242
   ||  ::::: |||||  ::::: |||||  ::::: |||||  ::::: |||||
120884 GGCCCTCGCTTCTTACATCTTCGACATGTTGGGCATTTCTGTGGTTTCATC 120933

243 LeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrIleTh 259
   |||||  |||||  |||  |||||  |||  |||||  |||  |||||
120934 GCATGGATGTTATTTGGTTAGCAGAACGCTACGCGATCATCCAGGATCTC 120983

259 rProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGln..... 273
   :  |||||  |||||  |||||  |||||  ::::: |||
120984 AGAATCGGCGGAGAGTATATCGAAGAGTCTACAGGTTACAGGCTAA 121033

273 ..... 273

121034 TAAATCAGGATCTAGCGGAGCGCGAGGAAGAGGAGGACGATGAAGTG 121083

274 .....LeuSerSerGlnLysSerValProTrpIleProMetLeuLys 287
   ::::: |||:::  ::::: |||:::  ::::: |||:::
121084 AGTCTGGCGGCGCGCGGAGGAACCGATACCCCTGTGCTCGCTCAC 121133

287 sSerLeuProLeuTrpAlaIleValAlaHisPheSerTyrAsnTrp 304
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
121134 ATCCGATCTCTGTGGGCCATCTTGTGACCGAATCGCGCCAGGATGGG 121183

304 hrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGluValLeu 320
   ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
121184 CCTTCTACACGACGATACCCGAGTGCACCTACATGAGCAACATCCTA 121233

321 ArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLeuG 337
   ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
121234 CACTTTGACATCCAGTCAATGCTGTGCTCAATGCGGTGCGGTACTTAAC 121283

337 yCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArg 354
   |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||
121284 CTCCTGGTCTCGGGCATTTGCTGCTCGCCCTCGCGGATTTGGATGCTAG 121333

354 laArgTrpAsnPheSerThrLeuTrpValArgValPheSerLeuIle 370

```

```

||||  ::|||  |||  :::::||||:  ::::: |||:::  ::::: |||:::
121334 CCAGACGCTACATATCGCTGCTGAACTCGTATAAGTTGTGGAACACGGTG 121383

371 GlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGlyCy 387
   ::  ::::: |||||  ::::: |||:::  ::::: |||:::  ::::: |||:::
121384 GCCTCGCGTGGTGCATCATCGGCTGATTGGCATCATCTATGTGGGCTG 121433

387 sAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyG 404
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
121434 CGATTGGGTGGGTACCTTTATGTGGCGCGGCTGGGCTCTTTTCGGCG 121483

404 lyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSer 420
   ||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
121484 CGCGCGCTATGCTGGCAACCATGATGATCATAGCGCTCAGTCCACGA 121533

421 TyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProG 437
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
121534 TATCAGGACCATGTTATGGCATCACCAATTCGGCGGCAATATCTGTGG 121583

437 yMetIleGlyPro...IleIleAlaArgSerLeuThrProGluAsnThrI 453
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
121584 CTTCTGGCTCCGTATGTCATCGGTCTTAATCATCAATCATCGCGAGACTC 121633

453 leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
   ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
121634 TGACACAGTGCATCTGTGCTTCTGGCTGGCGGCGGCTTGAATATAGCC 121683

470 GlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAsnTrpAl 486
   |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
121684 GGTAACTTCATCTACCTGATCTGCCAGCGCGGAGGCAAGCTGCTC 121733

486 a 486
   :
121734 G 121734

seq_name: gb_in:AE003491
seq_documentation_block:
LOCUS AE003491
DEFINITION Drosophila melanogaster genomic scaffold 142000013386053 section 8
of 30, complete sequence.
ACCESSION AE003491 AE002593
VERSION AE003491.2 GI:10728219
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 323461)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J., Basu,A.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,

```


Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattel,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarriy,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstein,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 333461)
Adams,M.D., Ceiniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7292821.
Location/Qualifiers
1. .323461
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
Join<15214. .15403,15749. .16006,16069. .16312,17722. .18091)
/gene="fne"
/note="CG4396"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
/protein_id="AAF48215.2"
/db_xref="GI:10728220"
/translation="MTNADIVKNGSANGSDGSDNDESTNLIIVLPOTWQEMRS
LFSSIGLESCKLVKDVSGSLGSGYGVNIVRAEDAERAVNTLGLRQNKVIKVSYA
RPSSEIKGANLYVSLPKNLSPDLQEGMFASFGRKIITSRLCDNISGLSGVGIRF
DORNEAERAIQELNGTKPKGYAEPITVKFANNPSAKAQIAPPLTAYLTPOAAATR
RLAGALPSAGRIRYPLAGDLLANSILPGNMTGSGWCIFVYNLAPTEENVLWOLFG
PEGAVOSVKVIRDLQTSCKGFGFTMYNDRAVVAIQSLNGYITLGNRVLQVSFTNK
TKTT"
complement(join(<23860. .24239,24538. .25184,25286. .25377,
25443. .25533,26089. .26389,26921. .>27095))
/gene="CG4395"
/product="CT4121"
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
complement(<23860. .>27095)
/gene="CG4395"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
complement(join(23860. .24239,24538. .25184,25286. .25377,
25443. .25533,26089. .26389,26921. .27095))
/gene="CG4395"
/note="CG4395 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"

/protein_id="AAF48216.1"
/db_xref="GI:7292823"
/translation="MATTSSDSESONVDVASQAQODNLRIFLKHLIAECVRYQNV
TVDDPSFSLGATDYDSDLPENFPVPRYLENAAMNEGVIDMRNVDEELAELMA
TVSATMATNOKENRILFCPLNFDGYLCWPRTAGTVLSQYCDFVEGPNRKLAKHTC
LNGSWYRHPVSNQTSNTYNCVDYEDLEFROFINELYKGYALSALLIILIFLG
FKSLRTRIRIHVHLFASLACTCVAWILMYRLVVERSETIAENPVCEKTSIENNIA
LOVARQLYKTNITINQFQICGLHLVHVFMELNVFMFCEGLHHLHLVIVVSYPS
GNLAICTLSLWATSVHGRCDFQLPTWFTPSNLSSEKSHRHLHLVIVVSYPS
GMSVOILTPLLSIQVFKDITVMRWFIVISWFSPIPTAIIVFLINLVIRVLRKLH
POSAOPALAIRKAVRATILVPLFGLQHLFPYRPDAGTOLDHFQMLSVLSVLSQ
FVVSFLCFANHDVFEATRILNKLPLSVTPPPAGSNWGMATTPSPSELGY"
Join(43393. .43627,43701. .43859,44240. .45178,45248. .46407,
47115. .>47153)
/gene="CG15721"
/product="CT35951"
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
<43393. .>47153
/gene="CG15721"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
Join(43393. .43627,43701. .43859,44240. .45178,45248. .46407,
47115. .>47153)
/gene="CG15721"
/note="CG15721 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
/protein_id="AAF48217.1"
/db_xref="GI:7292824"
/translation="MTAFLVSGIFALLAGARWDVGVPYPAPHYEQOQDLDLCP
SGPESLVACAGAPQFPSPCEVQRFALGTGINWEIHNRCETLEVCPCNCOQY
NPVCGKYDTRNFRSECLQVKRTGHPWRKHGDCESRYPPVQSPRVQNYPIYA
INPYQPNVNGRGAYSGQPGYQSGQPGQPGQPGQPGQPGQPGQPGQPGQPGQPGQ
OPYQIWSNLEPTEYAGTSKPYPNPTVETPTTSTASTTTSTTTTSTTTTSTTTT
DKITITPOGSDSETAASVTPNPRNIKNAVIEDTPKPSSTTTGQPESTESTSES
SVTTPLKKIITTAPOHATTARSPKPSLSINAVQVPEQAEKTSITEPTTYKSFN
SYNSASYTTTEAPVETDTPVDEIIRKINKVQISAENTTEKADETTOSTTLYNSTA
APTNTNTQLPDKVVKINAVEEKEIDEETKANGYETPVYSTEASKDTEITTK
SLDKVVKINAVEEKEIDEETKANGYETPVYSTEASKDTEITTK
LEKIVKNAIEESTNEKQDQETATVSSETSSSVYQTPDYKSVTPKNDVTQAP
STEASKDDETTKKSLDKVVKINAVEEKEDETTOKAETYSYPTYSASSVYKVPYAS
TEAPNDDETTKMLDKLVKINKAQISTEKAVNETTTTEPTITTKAPTTPSPQTYTSS
SVYKTSSETPDDEDEESTANPLNKSINKINAVSDPETIQTITATTPPETTKGSF
HGVKGEIPDADSVAKSAKTESAEIPEIVKTESRMIGYRDMQLEM
complement(join(<47450. .48160,48223. .48387,48516. .48897))
/gene="CG12716"
/product="CT35950"
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
complement(<47450. .>48897)
/gene="CG12716"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
complement(join(47703. .48160,48223. .48387,48516. .48897))
/gene="CG12716"
/note="CG12716 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
/protein_id="AAF48218.1"
/db_xref="GI:7292825"
/translation="MQTQQRKVSNSRAFTMILILQVLGILGRGVLAOKSSPIANWH
MADPHYTSQYAKILGEAGLQDADDKEHKLGLPMQIQYDVQPNQPGVPCATNG


```

||| ||||||||||||||||||||||||||| :||||:::
CCGAGATGCACGTACTCTTAATTGGCGCATTTCAAATGGTGTGCCAGCCT 1826
```

eProGlyWetIleGlyProlIeAlAArgSerLeuThrProGluAsnTr 452
1827 GTCGGGAATGGTCGCCCCGATCAATGTGTGGTGCAATGACGAAGAACAAGT 1876

hrlleGlyGIUtrGrInThValPhcCysIleAlAAIAlaIleASnVal 468
:: |::::| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1877 CCCGTGAAGAATGCGCATATGCTCTTCCTCATCGTCGCACTGGTCCCACAT 1926

pheCIyAlalIPhepheThrLeuPheAlaLysGlyGlyValGlraSnTr 485
||::|:|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1927 GGTCGAGTCATAATTTATGCACTATTTTCCTCAGGAGAGAACAACCTTG 1976

485 pALA 486
|||
1977 GC CA 1980

name: gb_pr:AB032435

<documentation_block>
S AB032435 3946 bp mRNA linear
NITION Homo sapiens DNPI mRNA for differentiation-associated
inorganic phosphate cotransporter, complete cds.
SSION AB032435
ION AB032435 GI:7328922
ORDS differentiation-associated Na-dependent inorganic phos-
phate transporter.

CE Homo sapiens brain thalamus cDNA to mRNa..
GANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; I
Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

(sites)
Ahara.Y., Mashima,H., Onda,H., Hisano,S., Kasuya.H.,
Yamada.S., Tomura,H., Yamada,Y., Inoue,I., Kojima.I.,
Molecular cloning of a novel brain-type Na(+)-depend-
ent phosphate cotransporter
J. Neurochem. 74 (6), 2622-2625 (2000)

URNAL 20281869
DLINE
RENCE 2 (bases 1 to 3946)
THORS Takeda,J. and Onda,H.
Direct Submission
Submitted (15-SEP-1999) Jun Takeda, Gunma University
Molecular and Cellular Regulation, Laboratory of Mol-
Genetics,Department of Cell Biology, 3-39-15, Showa-m-
Gunma 371-8512, Japan (E-mail:j.takeda@akagi.sbg.gunma.ac.jp)
Tel.:81-27-2220-8830, Fax:81-27-2220-8889)

URES location/Qualifiers
source 1 . 3946
organism="Homo sapiens"
db_xref="taxon:9606"
tissue_type="brain thalamus"
note="vector:pGEN3zf(+) ; site1,2:EcoRI"
414 . 2162
gene /gene="DNPI"
414 . 2162
cds /gene="DNPI"

/codon_start=1
product="differentiation-associated Na-depen-
dent inorganic phosphate cotransporter"
protein_id="BA092874.1"
db_xref="GI:7328923"
translation="MESVKQRLAPKCKGLKNFACKSLGSQIVRR
EDKGLEPPERKAJCDICTDCLPLPRRYITAIMSGLGFCISFIG
TLHRQGVIKKAKFNPDPTVMETVHTTFIMFGIIYTQPVGWI
LTSTNLMLPSAVRHLYGVCFVRILQLGLEVTYPACHGWTS
PCGSYAGVAIAMPLAGILVOYTWGSEVFVVYGSGFMWMFWEL
ERRRVIESIGESANLLGAMEKFKPWRKEFTSMPEYAILVAN
AFTEVFGEFSIKVGMLSAVPHLMWTIIVPGQTADFLRSKO
FGMEATLLLWGVSYSTRGVALSFVLAVGFSGFAISGNVNHL
GYGTLDSMWVCPIILVGAMTKNSREWEQVYFIALLAHVHYGGVII
EWTSBEKCGFHDELDDELDTGITONITYGTITKSYYAGTOAIN

seq_documentation_block:

inorganic ph

KEYWORDS differentiated

ORGANISM Homo sapiens

REFERENCE 1 (sites)

TITLE Molecular cl

MEDLINE	20281869
REFERENCE	2 (bases 1

JOURNAL OF THE DIRECT SUBMIT

Gunma 371-855

1. source

$$/t_i/$$

/ge

/co/

9p/
1d/

EDG
TTH

ECO
FER

GVG

LEI

```

GEVQDSHSYKDRVDYS"
polyA_signal 3895..3900
polyA_site 3921
BASE COUNT 1238 a 735 c 791 g 1182 t
ORIGIN
alignment_scores:
  Quality: 963.50      Length: 492
  Ratio: 2.737        Gaps: 9
  Percent Similarity: 71.545  Percent Identity: 41.057
alignment_block:
US-09-776-865-4 x AB032435 ..
Align seg 1/1 to: AB032435 from: 1 to: 3946
13 GlyGluGlu.....GlySerAspArgThrProLeuLeuGlnAr 25
|||||
506 GGAGAAGAAGCAAGACACCGGGGAGACAATCGAGCTGACGGAGGATGGGA 555
25 gAlaProArgAla.....Glu..ProAlaProValCys....CysSerAla 38
|||||
556 AGCCCTAGAGGTGCCCGAGAGAGAGCGCGCTGTGCGACTGCACCGTGC 605
39 .....ArgTyrAspLeuAlaPheLeuSerPhePheGlyPh 50
|||||
606 TTCGCCCTGCCCGCCCTACATTTATGCCCATCATGAGCGCCCTGGGCTT 655
50 ePheValLeuTyrSerLeuArgValAlaSerValAlaLeuValAspM 67
| : : : : :
856 CTGCATCTCCTTCGTATCCGCTCAACCTGGCGGTGGCCATTGTGGACA 705
67 etValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCys 83
| : : : : :
706 TGGTCAACACAGACACCATCCACCGCGGGGCAAGGTCAATCAAGGAGAAA 755
84 AlaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLy 100
|||
756 GCC .....AA 760
100 sTyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheT 117
| : : : : :
761 ATTCAACTGGGACCGGAAACCGTGGGATGATCCACGGTTCTCTTTT 810
117 yrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSer 133
| : : : : :
811 GGGGCTACATCATCAGTACAGATTCGGGAGGCTACATCGCGCTCTCGGCTG 860
134 GlyGlyLysLeuLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheTh 150
| : : : : :
861 GCACCCAAACAGGGTTTCGGAGCTGCCATACCTTCTTACCTACCCCTAAA 910
150 rLeuPheThrProLeuAlaAspPheGlyValIcIyAlaLeuValAlaL 167
| : : : : :
911 TATGCTAATTCATCAGCAGCAGCAGTGATGATGATGATGATGATGATGATG 960
167 euArgAlaLeuGluGlyLeuGlyGlyValThrTyrProAlaMetHis 183
| : : : : :
961 TCAGAACTACGACGGGACTTGTTCAGGGGTGTGACCTACCCAGCATGTCT 1010
184 AlaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSe 200
| : : : : :
1011 GGGATATGGACCAATGGGCCCCACCTCTAGAGAGGATAGACTGGCAAC 1060
200 rIleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeus 217
| : : : : :
1061 CACCTCTTTTGTGTTCTTCATGCGGAGCTGTGATTGCAATGCTTTAG 1110
217 erGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePhe 233
| : : : : :
1111 CTGGCATCTTGTGCTGACATACACTGGCTGGTCTTCAGTGTCTTATGTCTAC 1160

```

```

234 GlyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAs 250
|||
1161 GGAAGCITTTGAATGGCTCGTACATGTTTGGCTTTGGTGTCTTATGA 1210
250 pThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleL 267
| : : : : :
1211 AAGTCCTGCAAGCATCTCTACTATTACAGATGAAGACGCTAGGTACATAG 1260
267 euSerSerLeu.....LysAsnGlnLeuSerSerGlnLys..... 278
| : : : : :
1361 AAGAAACATTTGGAGAGAGTCAAAATCTTTAGTGTGCAATGGAAAAATTC 1310
279 SerValProTrpIleProMetLeuLysSerLeuProLeuTrpAlaIleVa 295
| : : : : :
1311 AAGACTCCATGGGAAGTTTTTCATCCATCCAGTCAGTCATGCAATAAT 1360
295 lValAlaHisPheSerTyrAsnTrpThrPheTyrThrLeuLeuThrLeuL 312
| : : : : :
1361 TGTTCGAAACTCTCTGCAAGAGCTGGACTTTTATTATTATTGCTTATTAGTC 1410
312 euProThrTyrMetLysGluValLeuArgPheAsnIleGlnGluAsnGly 328
| : : : : :
1411 AGCCAGCATATTTTGAGGAGTCTTTGGATTGAAATTAGCAAGGTGGT 1460
329 PheLeuSerAlaValProTyrLeuGlyCysTrpLeuCysMetIleLeuSe 345
| : : : : :
1461 ATGCTATCTGCTGGCCACACTTAGTAATGACAATTTATGTCCTATTTGG 1510
345 rGlyGlnAlaAlaAspAsnLeuArgAlaArGrpAsnPheSerThrLeuT 362
| : : : : :
1511 GGGCAAAATTCAGATTTTCTAAGAAGCAAGCAGATCTCTTCAACCTACGA 1560
362 rpValArgArgValPheSerLeuIleGlyMetIleGlyProAlaIlePhe 378
| : : : : :
1561 CAGTGAGAAAGATCATGAATTTGGTGGTGTGGCATGGAGCCACACTG 1610
379 LeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaValAlaPh 395
| : : : : :
1611 CTCTGCTGCTGGCTAT...TCTCATACTAGAGGGGTAGCAATCTCAT 1657
395 eLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGlyPheSerI 412
| : : : : :
1658 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1707
412 leAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeuLeuGlyIle 428
| : : : : :
1708 TTAACCACTTGGATATCGCTCCAGATATGCCAGTATCTTAAATGGGCT 1757
429 ThrAsnThrPheAlaThrIleProGlyMetIleGlyProIleIleAlaAr 445
| : : : : :
1758 TCGAATGGTGTGGCAGATGTCAGGAATGGTGTGCTCTATCATTTGTTGG 1807
445 gSerLeuThrProGluAsnThrIleGlyGluTrpGlnThrValPheCysI 462
| : : : : :
1808 TGCAATGCAAGAAGAAATAGTCAGCTGAAGAGTGGCAGTGTCTCTCTGA 1857
462 leAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrLeuPheAla 478
| : : : : :
1858 TCGCTGCCCTAGTCCACTATGTTGGAGTTATTTTATGCAATATTGGCC 1907
479 LysGlyGluValGlnAsnTrpAla 486
| : : : : :
1908 TCAGGAGAGAAACAACCCCTGGCA 1931

```


THIS PAGE BLANK (USPTO)

alignment_block:

US-09-776-865-4 x AAZ50876 ..

Align seg 1/1 to: AAZ50876 from: 1 to: 2844

```
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||
84 ATGAAGTCCCGGTTTCGGACTTAGCCCGGACGCGGAGGAGGGCTC 133
|||||
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProv 34
|||||
134 GGACCGCACACCGCTCCCTGCAGCGGCCCGCGGGCGGAACCGCTCCAG 183
|||||
34 aLysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
|||||
184 TATGCTGCTCTGCTGCTTACAACTAGACATTTTGTCTTTTGGTTTC 233
|||||
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
234 TTCGTTCTCTATTCAATACGGGTGAATCTGAGCGTTGCACCTAGTGGACAT 283
|||||
67 tValAspSerAsnThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
284 GGTGGATTCAAAACACAACTGCCAAAGATAATAGAACGCTCTACGAGTGTG 333
|||||
84 LaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
334 CAGAGCATCTGCTCCCAATAAAGTCTTCACAAACCAACGGGTAAAAAG 383
|||||
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
384 TACCGGTGGGATGCAGAAACTCAAGGATGGATTCTCGGATCTTTTCTA 433
|||||
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
434 TGGCTACATCATCACAAATTCCTGGAGGATATGTTGCCAGCAGAAAGTG 483
|||||
134 LyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
484 GGGGAACCTGTTGCTAGANTTCGGATCTTTGCTACAGCTATCTTCACC 533
|||||
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
534 CTGTTCACTCCCTCGCTGCAGATTTCGGAGTCGGAGCCCTTTGTGCAC 583
|||||
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMetHisA 184
|||||
584 CAGGGCCTAGAAAGGCTTAGGAGAGGGGTGCACATATCCAGCCATGCATG 633
|||||
184 LaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
634 CCATGTGCTCTCATGGGCTCCCTCTTGAAGAAGCAAGCTTCTGAGT 683
|||||
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
684 ATTTTCATATGCAGGACACACACTTGGGACAGTAGTCTCTCTCTCTTC 733
|||||
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||
734 TGGAGTAATTTGCTACTATATATGAAATGGACTTATGTCTTCTATTCTTTG 783
|||||
234 LyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
784 GCATGTGTGGAATCATCTGGTTATTTATGGATCTGCTTAGTGTAGTGAT 833
|||||
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
834 ACACCAGAACTCACAGACAATCACTCCCTATCGAAAAGGAGTATATCT 883
|||||
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
884 TTCATCATTAATAAATCAGTCTCTTCACAGAGTCAGTGCCTGGATAC 933
|||||
```

```
284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
|||||
934 CTATGCTGAATCACTGCCACTTTGGGCTATTGCTGTGCACATTTTCT 983
|||||
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||
984 TACAACTGGACTTTTATACCTTTGTTGACCTTATTTGCTTACATGAA 1033
|||||
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
1034 GGAAGTCTCAAGGTCAATATTAAGAGATGGGTTTTTATCTGACGTCC 1083
|||||
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||
1084 CTTATTAGGTGTTGTTATGATGATCCTGCGGTCAAGCTGCTGAC 1133
|||||
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1134 AATTAAGGGCAAGATGGAATTTTCAACTCTGTGGGTTCGAAGAGTTTT 1183
|||||
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1184 TAGCCTTATAGGGATGATTGGACCTGGCATATTCTGTTGCCGAGGAT 1233
|||||
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1234 TTATAGGCTGTGATTATCTTGGCTGTTCATTCCTTAACCATATCAACA 1283
|||||
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1284 ACCCTGGAGGCTTTTGTCTTCTGGATTTAGCATCAACCATCTGGACAT 1333
|||||
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaP 434
|||||
1334 TGCTCTCTCGTATGCTGTGTTATCTCTGGCATCACAATACCTTTGCCA 1383
|||||
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1384 CTATTCCTGGNATGATTGGGCCCAATCATGCCAGAAAGTCTTACCCCTGAG 1433
|||||
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1434 AACACTATTGGAGAAATGGCAAACTGTTTCTGTCATCGCTGCTATCAA 1483
|||||
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1484 TGTATTTGGTGCCATTTTCTTCACTATTTCGCCAAAAGGTGAAGTGCAA 1533
|||||
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||
1534 ACTGGGCATCAGTATCATCCACAGGACACAGAAAC 1568
|||||
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAD10326
seq_documentation_block:
ID AAD10326 standard; DNA; 2844 BP.
XX
AC AAD10326;
XX
DT 16-OCT-2001 (first entry)
XX
DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) DNA.
XX
KW Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;
KW cytostatic; vulnery; antiatherosclerotic; osteopathic; vasotropic;
KW prevention; attenuation; pathogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine; ds.
XX
OS Ovis sp.
XX
```

FH Key Location/Qualifiers
FT CDS 84..1571
FT /*tag= a
FT /product= "Sheep GBS toxin receptor protein (SP55)"
XX WO200156598-A2.
XX 09-AUG-2001.
XX 02-FEB-2001; 2001WO-US03662.
XX 02-FEB-2000; 2000US-0179870.
XX (UYVA-) UNIV VANDERBILT.
XX Hellerqvist CG;
XX WPI: 2001-488844/53.
XX P-PSDB; AAE06519.
XX Preventing or attenuating pathoangiogenic conditions e.g. cancer,
XX chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
XX administering group B beta-hemolytic Streptococci toxin receptor or its
XX fragment
XX Disclosure: Page 47-50; 52pp; English.
XX The present sequence is a DNA encoding sheep group B beta-haemolytic
XX Streptococci (GBS) toxin receptor protein, SP55. The present invention
XX relates to a method for preventing or attenuating a patho-angiogenic
XX condition in a mammal which comprises administering to the mammal one
XX or more GBS toxin receptors or their immunogenic fragments to induce
XX or maintain an immune response to one of GBS toxin receptors. The
XX method is useful for preventing or ameliorating pathoangiogenic
XX conditions such as cancer, scarring during wound healing, gliosis
XX during repair of nerve injury, chronic wounds, keloids, reperfusion
XX injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
XX psoriasis in mammals. The proteins of the invention are also used
XX as vaccines.
XX Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

alignment_scores:
Quality: 2617.00 Length: 495
Ratio: 5.287 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-776-865-4 x AAD10326 ..,
Align seg 1/1 to: AAD10326 from: 1 to: 2844
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlyLe 17
84 ATGAAGTCCCGGTTTCGGACTTAGCCCCGAGCGACGGGAGGGGCTC 133
17 rAspArgThrProLeuGlnArgAlaProArgAlaGluProAlaPro 34
134 GGACCGCACACCGCTCTGAGCGCGCGCGCGGGGGAACCGCTCCAG 183
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
184 TATGCTGCTCTGCTCGTTACAACCTAGCATTTTGTCTCTTTTGGTTTC 233
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAsp 67
234 TTCGTTCTCTATTATACGGGTGAATCTGAGCGTTTGCACCTAGTGGACAT 283
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCys 84
284 GGTGGATTCAACACACACTCCCAAGATAATAGAACGCTCTACGAGGTG 333

```
|||||
1234 TTATAGGCTGTGATTATTCCTTGGCTGTGATTCTTAACCATATCAACA 1283
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIle 417
|||||
1284 ACCCTGGAGGCTTTTGTCTCTCTGGATTTTASCATCAACCATCTGGACAT 1333
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaIat 434
|||||
1334 TGCCTCTCGTATGCTGTATCTCTGGGCATCACAAATACCTTTGCCA 1383
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1384 CTATTCCTGGAATGATGGGCCCATCATTCGCCAGAAGCTTACCCCTGAG 1433
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1434 AACACTATTGGAATGCAAACTGTTTCTGTCATCGCTGCTGCTATCAA 1483
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1484 TGTATTGGTGCCATTCTTTCACACATATTGCCCAAGGTGAAGTGCAAA 1533
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||
1534 ACTGGCCATCAGTGATCACCAAGSACACAGAAAC 1568
```

seq_name: /SDSI/gc9data/hold-geneseq/geneseqn-emb1/WA2001A.DAT:AAF55900

seq_documentation_block:

ID AAF55900 standard; DNA; 2512 BP.

AC AAF55900;

DT 18-APR-2001 (first entry)

DE Human AST coding sequence.

XX Human: AST; nootropic; immunotropic; gene therapy; Salla disease;

XX anion and sugar transporter; anion-cation symporter;

XX sialic acid transporter; ss.

OS Homo sapiens.

PN EP1069184-Al.

XX 17-JAN-2001.

XX 16-JUL-1999; 99EP-0202341.

XX 16-JUL-1999; 99EP-0202341.

XX (ALKU) AKZO NOBEL NV.

XX WPI; 2001-193090/20.

DR P-PSDB; AAB66967.

XX New human transporter gene implicated in Salla disease and lysosomal

PT sialic acid transport, useful in assays for identifying new drugs, or

PT diagnosing sialic acid transport defects related to mutations in the

PT transporter gene

XX Claim 2; Page 12-13; 20pp; English.

XX The present sequence is the coding sequence for human Anion and Sugar

CC transporter (AST) protein. AST has significant homology with several

CC members of the anion-cation symporter (ACS) family of transporters. AST

CC is implicated in Salla disease, and is useful in screening assays for

CC identifying new drugs. Compounds identified via AST screening is useful

CC for preparing a pharmaceutical suitable as an activator or inhibitor of a

CC sialic acid transporter protein. The pharmaceutical may be used in sialic

CC acid associated diseases and CNS/immune related disorders.

XX

SQ Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other;

alignment_scores:
Quality: 2329.00 Length: 495
Ratio: 4.903 Gaps: 0
Percent Similarity: 95.960 Percent Identity: 86.263

alignment_block:
US-09-776-865-4 x AAF55900 ..

Align seg 1/1 to: AAF55900 from: 1 to: 2512

```
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||
271 ATGAGGTCTCGGTTCGAGACCTGGCCCGGAACGATGGCGAGAGACAC 320
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
|||||
321 GGACCGCACCGCTCTTCTACGGGCGCCCGGCGGCGGAGCCGCTCCAG 370
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
|||||
371 TGTGCTGCTCTGCTCGTTACAACTTAGCAATTTTGGCTTTTGGTTTC 420
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
421 TTCAATTGTGATGATTACGTGTGAATCTGAGTGTTCGTTAGTGGATAT 470
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
471 GGTAGATTCAATACAACTTTAGAAGATAATAGAACTTCCAAGCGCTGTC 520
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
521 CAGAGCATTTCTGCCATATAAAGTTTCATATAATCAAGGGTAAGAAG 570
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
571 TACCAATGGGATGAGAACTCAAGGATGGATTCCTCGTTCCTTTTITTA 620
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
621 TGGCTACATCATCACACAGATTCTCTGGAGGATATGTTGCCAGCAAAATAG 670
134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
671 GGGGGAATAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGCTCACC 720
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
721 CTGTTCACTCCCAATGCTGAGATTTAGGAGTTGGACCACTCAATTGTACT 770
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHis 184
|||||
771 CAGAGCACTAGAAGGACTAGGAGAGGGTGTACATTTCCAGCCATGCATG 820
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
821 CCATGTGCTCTTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTTAGC 870
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
871 ATTTTCATATCAGGAGCACAGCTTGGGACAGAGTAATTTCTTCTCTCTTC 920
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||
921 TGGAAATAATTTGCTACTATATGAATTTGGATGCTTCTTCTACTTTTTC 970
234 lyIleValGlyIleIleThrPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
971 GTACTATTGGAATATTTGTTCTTTTGTGGATCTGTTAGTGTAGTGAC 1020
```

```
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLeu 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1021 ACACCACACAAACACAGAGATTTCCCATATGAAAGGAATACATTTCT 1070
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1071 TTCATCATTAAGAAATACAGTTTCTTCACAGAAAGTCAGTGGCTGGGTAC 1120
284 roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1121 CCATTTTAAATCCCTGCCACCTTGGGCTATGCTAGTTGCACACTTTTCT 1170
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetly 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1171 TACAACCTGGACTTTTATACATTTTATGACATTTATGCTTACTATATGAA 1220
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1221 GGAGATCCTTAAGTTCAATGTTTCAGAGAAATGGGTTTATCTTCATTCG 1270
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1271 CITATTAGGCTCTGGTTATGATGATCTCTCTGGTCAAGCTGCTGAC 1320
351 AsnLeuArgAlaAArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1321 AATTTAAGGGCAAAATGGAATTTTCAACTTTATGTGTTCGCAGAAATTT 1370
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1371 TAGCCTTATAGGAATGATTTGGACCTGCAGATTTCTCTGGTAGCTGGCT 1420
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1421 TCATTTGGCTGTGATTTCTTTGGCGTGTCTTCTTAACATATACACA 1470
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspI 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1471 ACACCTGGGAGGCTTTTGTCTTCTGATTTAGCATCAACCATCTGGATAT 1520
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1521 TGCTCTCTGATGCTGGTATCTCTCTGGGCATCACAAATACATTTGCCA 1570
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1571 CTATTTCCAGGAATGTTGGCCCGTCATTTGCTAAAGTCTGACCCCTGAT 1620
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1621 AACACTGTTGGAGAATGGCAACCGTGTCTATATTGCTGCTATTATA 1670
467 nValPheGlyAlaIlePhePheThrIleuPheAlaLysGlyGluValGlnA 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1671 TGTTTTGGTGCCATTTCTTTACACTATTCGCCAAAGGTGAAGTACAAA 1720
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1721 ACTGGGCTCTCAATGATCACTGATGACACAGAC 1755
```

```
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ50875
seq_documentation_block:
ID AAZ50875 standard; cDNA; 2602 BP.
XX
AC AAZ50875;
XX
DT
XX
XX 31-MAY-2000 (first entry)
XX Partial human GBS toxin receptor (HP55) cDNA.
XX Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KW pathological vascularisation; cancer metastases; angiogenesis;
```

```
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 58..1545
FT /*Tag= a
FT /product= "Partial human GBS toxin receptor"
XX
FN WO200005375-A1.
XX
PD 03-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16676.
XX
PR 22-JUL-1998; 98US-0093843.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI HELLERQVIST CG, Fu C;
XX
DR WPI; 2000-205377/18.
DR P-PSDB; AAY45087.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX useful for diagnosis and treatment of, e.g. pneumonia in neonates -
PS Claim 3; Page 77-80; 109pp; English.
XX
CC The present cDNA sequence encodes partial human GBS (group B beta
CC -haemolytic streptococci) toxin receptor (HP55). This sequence was cloned
CC by using human embryo lung cDNA library as template. Expression vectors
CC comprising this cDNA can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
SQ Sequence 2602 BP; 672 A; 576 C; 583 G; 771 T; 0 other;

alignment_scores:
Quality: 2329.00 Length: 495
Ratio: 4.903 Gaps: 0
Percent Similarity: 95.960 Percent Identity: 86.263

alignment_block:
US-09-776-865-4 x AAZ50875 ..
Align seg 1/1 to: AAZ50875 from: 1 to: 2602

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 ATGAGGTCTCCGGTTCGAGACCTGGCCCGGAACGATGGCGAGGAGACAC 107
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 GGACCGCACGCGCTTCTCTACCGGGCGCCCGCCGCGGCGAAGCGCTCCAG 157
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 TGTGCTGCTCTGCTCTGCTTACACTTAGCAATTTTGGCCTTTTGGTTTC 207
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
```


XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
 PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
 XX
 PS Claim 3; Page 89-93; 109pp; English.
 XX
 CC The present cDNA sequence encodes full length human GBS (group B beta
 CC -haemolytic streptococci) toxin receptor (HP59): This sequence was cloned
 CC by using human embryo lung cDNA library as template. Expression vectors
 CC comprising this cDNA can be transformed into host cells to express GBS
 CC toxin receptor and its fragments. Detecting the receptor in tissues is
 CC used to diagnose pathological vascularisation, e.g. for detecting cancer
 CC metastases. GBS toxin receptors are useful for treating conditions
 CC associated with pathological angiogenesis or neovascularisation
 CC (specifically cancer, reperfusion injury, scarring during wound healing,
 CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
 CC neural injury), and to raise specific antibodies used for treating early
 CC onset disease. Inhibitors of this receptor are useful for treating
 CC pathological or hypoxia-induced endothelial cell proliferation and
 CC migration.
 XX
 SQ Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

alignment_scores:
 Quality: 2329.00 Length: 495
 Ratio: 4.903 Gaps: 0
 Percent Similarity: 95.960 Percent Identity: 86.263

alignment_block:
 US-09-776-865-4 x AAZ50879 ..

Align seg 1/1 to: AAZ50879 from: 1 to: 2930

```

1  MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 ATGAGTCTCCGGTTGAGACCTGGCCCGGAGACGATGGCGAGAGAC 435
17  rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProv 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 GGACCCGACGCGCTTCTACCGGGCGCCCGCCAGCGCGCGCTCCAG 485
34  aLcysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 TGTGCTGCTGCTGCTTCACTTACCAATTTGGCCCTTTTGGTTC 535
51  PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspHe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 TTCATTGTGTATGATACGTTACGTGAATCTGAGTGTTCGCTTGGATAT 585
67  tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
586 GGTAGATTCAAAATACAACTTTAGAGATATAGAACTTCAAGCGCGTGC 635
84  laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
636 CAGAGCATTTCTGCTCCCATAAAGTTTCATCATATCAAAACGGGTAAGA 685
101  TyrArgTyrAspAlaGluThrGlnGlyTyrPheLeuGlySerPhePheTy 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
686 TACCAATGGGATGCGAAGACTCAAGGATGGATTTCTCGGTTCCTTTT 735
117  rGlyTyrIleLeuThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
736 TGGCTACATCATCACACAGATTCTCGGAGATATGTTGCCAGCAAAATAG 785
134  lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
786 GGGGGAATAATGCTGCTAGGATTTGGGATCTTGGCCTGCTGCTCACC 835
151  LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
836 CTGTTCACTCCCATTCGTCAGATTTAGGAGTTGGACCACCTCATTTACT 885

```

```

167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethisA 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 CAGACACTAGAAAGGACTAGGAGAGGCTGTACATTTCCAGCCATGCATG 935
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 CCATGTGGTCTCTTGGGCTCCCTCTTGAAGAAGCAAACTCTTATGAGC 985
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 ATTTCGTATGCAGAGACACAGCTTGGACAGTAATTTCTCTCTCTTTC 1035
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 TGAATAAATTTGCTACTATATGAATGGACTTATGTCTTCTACTTTTTC 1085
234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 GTACTATTGGAATATTTTGGTTCTTTTGGATCTGGTTAGTTAGTGAC 1135
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 ACACACAAAAACACAGAGAAATTTCCCATATGAAAGGAATACATTCT 1185
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 TTCATCATTAAGAAATCAGCTTTCTTACAGAGAATCAGTCCCGTGGTAC 1235
284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1236 CCATTTAAATCCCTGCCACTTTGGCTATCGTAGTTGCACACTTTTCT 1285
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1286 TACAACCTGGACTTTTATACCTTATTGACATATATGCCTACTTATATGAA 1335
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1336 GGAGATCTTAAGTTCATCAATGTTCAAGAGAAATGGTTTATCTTCATTGC 1385
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1386 CTATTATTAGGCTCTTGGTTATGATGATCTGCTGCTCAAGCTGCTGAC 1435
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1436 AATTTAAGGCAAAATGGAATTTTCAACTTATGTGTTCCGCAAAATTT 1485
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1486 TAGCCTTATAGGAATGATTGGACCTGCAGTATTCTCTGGTAGCTGGCT 1535
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1536 TCATTGGCTGTGATTATTCTTGGCCGCTGCTTCTTAAGTATATCAACA 1585
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1586 ACACCTGGAGGCTTTTGTCTCTCTGGATTAGCATCAACCATCTGGATAT 1635
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaI 434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1636 TGCTCCTCGTATGCTGGTATCTCTCTGGGCGCATCACAAATACATTGGCA 1685
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1686 CTATTCCAGGAATGGTTGGCCCGCTCATTTGCTAAAAAGTCTGACCCCTG 1735
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1736 AACACGTGTTGGAGAAATGGCAACCGGTTCCTATATTGCTGCTGCTATTAA 1785

```


467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
 |||||TGTGGTGGCCATTTCTTACACTATTCGCCAAAGGTGAAGTACAAA 1835
 1786 TGTGGTGGTGGCCATTTCTTACACTATTCGCCAAAGGTGAAGTACAAA 1835
 484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
 |||||TGTGGTGGCCATTTCTTACACTATTCGCCAAAGGTGAAGTACAAA 1835
 1836 ACTGGGCTCTCAATGATCACCATTGGACACAGACAC 1870

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT: AAD10325

seq_documentation_block:

ID AAD10325 standard; DNA; 2930 BP.

XX
 AC AAD10325;

XX
 DT 16-OCT-2001 (first entry)

XX
 DE Human group B beta-haemolytic Streptococci toxin receptor (HP59) DNA.

XX
 KW Human; group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;
 KW cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic;
 KW prevention; attenuation; pathoangiogenic condition; cancer; scar; injury;
 KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
 KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
 KW vaccine; ds.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT CDS 263..1873

XX
 FT /*tag= a
 FT /product= "Human GBS toxin receptor protein (HP59) "

XX
 PN WO200156598-A2.

XX
 PD 09-AUG-2001.

XX
 PF 02-FEB-2001; 2001WO-US03662.

XX
 PR 02-FEB-2000; 2000US-0179870.

XX
 PA (UYVA-) UNIV VANDERBILT.

XX
 PI Hellergvist CG;

XX
 DR WPI: 2001-488844/53.

XX
 DR P-PSDB; AAE06518.

XX
 PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
 PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
 PT administering group B beta-hemolytic Streptococci toxin receptor or its
 PT fragment -

XX
 PS Disclosure; Page 41-44; 52pp; English.

XX
 CC The present sequence is a DNA encoding full length human group B beta-
 CC haemolytic Streptococci (GBS) toxin receptor protein, HP59. The present
 CC invention relates to a method for preventing or attenuating a patho-
 CC angiogenic condition in a mammal which comprises administering to the
 CC mammal one or more GBS toxin receptors or their immunogenic fragments
 CC to induce or maintain an immune response to one of GBS toxin receptors.
 CC The method is useful for preventing or ameliorating pathoangiogenic
 CC conditions such as cancer, scarring during wound healing, gliosis
 CC during repair of nerve injury, chronic wounds, keloids, reperfusion
 CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
 CC psoriasis in mammals. The proteins of the invention are also used
 CC as vaccines.

XX
 SQ Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

alignment_scores:

Quality: 2329.00

Length: 495

Ratio: 4.903 Gaps: 0
 Percent Similarity: 95.960 Percent Identity: 86.263
 alignment_block:
 US-09-776-865-4 x AAD10325 ..
 Align seg 1/1 to: AAD10325 from: 1 to: 2930
 1 MetLysSerProValSerAspLeuAlaProSerAspGlycylGluGlySe 17
 386 ATGAGGTCTCCGGTTCGAGACCTGGCCCGGACGATGGCGGAGGAGAC 435
 17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaPro 34
 436 GGACCGCAGCGCTCTTCTACCGGGCGCCACCGGGCGAAGCGCTCCAG 485
 34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
 486 TGTGCTGCTCTGCTCGCTTACAACTTAGCAATTTTGGCCTTTTGGTTTC 535
 51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
 536 TTCATTGTGTATGATACGTGTAATCTGAGTGTGCGTTAGTGGATAT 585
 67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
 586 GGTAGATTCAATACAACTTTAGAGATAATAGAACTTCCAAGCGGTGC 635
 84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
 636 CAGAGCATTCCTGCTCCCATAAAAGTTCATCATATAATCAACGGGTAAGA 685
 101 TyrArgTrpAspAlaGluThrGlnGlyTyrIleLeuGlySerPhePheTy 117
 686 TACCAATGGGATGCAGAACTCAAGGATGGATTCTCGGTTCTCTTTTAA 735
 117 rGlyTyrIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
 736 TGGCTACATCATCACAGATTCTTGAGGATATGTTGCCAGCAAAATAG 785
 134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
 786 GGGGGAATAATGCTGCTAGGATTGGGATCCTTGGCCTGCTCTCCTCACC 835
 151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
 836 CTGTTTCTCTCCCATTCGTCAGATTAGGAGTTGGACCATCTCAATTGTA 885
 167 uArgAlaLeuGlyGlyLeuGlyGlyValThrTyrProAlaMethHisA 184
 886 CAGAGCAGTAGAAGGAGGACTAGGAGGGGTGTACATTTCCAGCCATGCA 935
 184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
 936 CCATGTGGTCTTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTAGC 985
 201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
 986 ATTTCGTATGACGAGGACACAGCTTGGGACAGTAATTTCTTCTCTTTT 1035
 217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
 1036 TGGAAATAATTTGCTACTATATGATGGACTTATGCTTCTACTCTTTT 1085
 234 lyTleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
 1086 GTACTATTGGAATATTGGTTTCTTTTGGGATCTGGTTAGTTAGTGAC 1135
 251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
 1136 ACACCAACAAAACACAGAGAAATTTCCCATTTATGAAAGGAATACATTCT 1185
 267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284


```
51  PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 TTCAATTGTTGATGCATTACGTTGAATCTGAGTGTTCGTTAGTGGATAT 200
67  tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 GGTAGATTCAAAATCAACTTTTAGAAGATAATAGAACTTCCAAGCGGTGC 250
84  LaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 CAGAGCATTCCTGCCCAATAAAGTTTCATCATATAATCAACGGGTAAAG 300
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPheTy 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 TACCAATGGGATGCAGAACTCAAGGATGGATTCTCGSTTCTTTTAA 350
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerg 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 TGGCTACATCATCACAGATTCTCTGGAGGATATGTTGCCAGCAAAATAG 400
134 LyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 GGGGAAATGCTGCTAGGATTTGGGATCTCTGGCACTGCTGCTCCTCACC 450
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 CTGTTCACTCCCATTTGTCAGATTAGGAGTTGGACCTCATTTGTAAT 500
167 uArgAlaLeuGluGlyLeuGlyGluValThrTyrProIleMethHisA 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 CAGAGCACTAGAAAGCTAGAGAGGGTGTACATTTCCAGCCATGCATG 550
184 laMetTrpSerTrpAlaProLeuGluArgSerLysLeuLeuSer 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 CCATGTGCTCTCTTGGGCTCCCCCTCTTGAAGAAGCAAACTCTTAGC 600
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 ATTTTCATATGCAGAGCACAGCTTGGGACAGTAATTTCTCTCTCTTTC 650
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 TGGAAATAATTGCTACTATATGAATTGGACTTATGTCTTCTACTTTTTTG 700
234 LyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 GTACTATTGGAAATATTTGGTTCTTTTGGGACTCGTTAGTTAGTGAC 750
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 ACACCACAAAAACACAGAGAAATTTCCCATTTATGAAAAGGAATACATCT 800
267 uSerSerLeuLysAsnGlnIleuSerSerGlnLysSerValProTrpIleP 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 TTCACTATTAGAATAACAGCTTTCTTCACAGAGTCAGGCGCGGTGAC 850
284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 CCATTTTAAATCCCTGCCACTTTGGGCTATCGTAGTTGCACACTTTTCT 900
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 TACAACCTGGACTTTTATACTTTTATTCACATTAATTGCCTACTTATAGAA 950
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 GGAGATCCTCAAGGTCAATGTTCAAGAGAATGGGTTTTTATCTTCATTGC 1000
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1001 CTTATTTAGGCTCTGGTTATGATGATCTGCTGCTGCTGCTGCTGCTGAC 1050
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgValPh 367
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1051 AATTTAAGGGCAAAATGGAATTTTCAACTTTATGTCTTCGCAGAAATTTT 1100
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1101 TAGCCTTATAGGAATGATTGGACCTGCAGATTTCCTGGTAGCTGGCT 1150
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1151 TCATTGGCTGTGATTATTTCTTGGCCGTTGCTTCTTAACATATATCAACA 1200
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1201 ACACCTGGAGGCTTTTCTCTCTGGATTAGCATCAACCATCTGGATAT 1250
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1251 TGTCTCTCTGATGCTGGTATCTCTCTGGGCATCACAAATACATTTGCCA 1300
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1301 CTATTTCCAGGAATGGTTGGGCCGCTCATTTGCTAAAGCTTGACCCCTGAT 1350
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1351 AACACTGTTGGAGAAATGGCAACCGTGTCTATATTGCTGCTGCTATTA 1400
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1401 TGTTTTTGGTGCCATTTTCTTTACACTATTTCGCCAAAGSTGAAGTACAAA 1450
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1451 ACTGGGCTCTCATGATCACCATGGACACAGACAC 1485
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AAK94876
seq_documentation_block:
ID AAK94876 standard; cDNA; 2712 BP.
XX
AC AAK94876;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 4068.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
(HELI-) HELIX RES INST.
PA
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
DR P-PSDB; AAM93914.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
PS Claim 8; SEQ ID NO 4068; 1380pp + sequence listing; English.
XX
```



```
467 AsnValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGI 483
|||||
1583 AATGTTTTGGTCATTTCTTTTACACTATTTCGCAAGGTGAAGTACA 1632
|||||
483 nAsnTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||
1633 AAACGGGCTCTCATGATCACCATTGACACAGACAC 1669

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ50880
seq_documentation_block:
ID AAZ50880 standard; cDNA; 1485 BP.
XX
AC AAZ50880;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human/Sheep consensus GBS toxin receptor cDNA sequence-1.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX
OS Homo sapiens.
XX
OS Ovis sp.
XX
FH Key
FT mat_peptide
FT 1..1485
FT Location/Qualifiers
FT *tag= a
FT /product= "Human/sheep consensus GBS toxin receptor
FT /protein sequence"
FT /transl_except= (pos:4..6; aa:Xaa)
FT /transl_except= (pos:7..9; aa:Xaa)
FT /transl_except= (pos:16..18; aa:Xaa)
FT /transl_except= (pos:22..24; aa:Xaa)
FT /transl_except= (pos:28..30; aa:Xaa)
FT /transl_except= (pos:31..33; aa:Xaa)
FT /transl_except= (pos:34..36; aa:Xaa)
FT /transl_except= (pos:46..48; aa:Xaa)
FT /transl_except= (pos:49..51; aa:Xaa)
FT /transl_except= (pos:58..60; aa:Xaa)
FT /transl_except= (pos:61..63; aa:Xaa)
FT /transl_except= (pos:64..66; aa:Xaa)
FT /transl_except= (pos:67..69; aa:Xaa)
FT /transl_except= (pos:70..72; aa:Xaa)
FT /transl_except= (pos:73..75; aa:Xaa)
FT /transl_except= (pos:79..81; aa:Xaa)
FT /transl_except= (pos:85..87; aa:Xaa)
FT /transl_except= (pos:91..93; aa:Xaa)
FT /transl_except= (pos:100..102; aa:Xaa)
FT /transl_except= (pos:124..126; aa:Xaa)
FT /transl_except= (pos:130..132; aa:Xaa)
FT /transl_except= (pos:136..138; aa:Xaa)
FT /transl_except= (pos:154..156; aa:Xaa)
FT /transl_except= (pos:157..159; aa:Xaa)
FT /transl_except= (pos:163..165; aa:Xaa)
FT /transl_except= (pos:169..171; aa:Xaa)
FT /transl_except= (pos:181..183; aa:Xaa)
FT /transl_except= (pos:187..189; aa:Xaa)
FT /transl_except= (pos:190..192; aa:Xaa)
FT /transl_except= (pos:196..198; aa:Xaa)
FT /transl_except= (pos:202..204; aa:Xaa)
FT /transl_except= (pos:211..213; aa:Xaa)
FT /transl_except= (pos:220..222; aa:Xaa)
FT /transl_except= (pos:223..225; aa:Xaa)
FT /transl_except= (pos:235..237; aa:Xaa)
FT /transl_except= (pos:241..243; aa:Xaa)
FT /transl_except= (pos:244..246; aa:Xaa)
FT /transl_except= (pos:250..252; aa:Xaa)
FT /transl_except= (pos:277..279; aa:Xaa)
FT
FT /transl_except= (pos:280..282; aa:Xaa)
FT /transl_except= (pos:283..285; aa:Xaa)
FT /transl_except= (pos:295..297; aa:Xaa)
FT /transl_except= (pos:304..306; aa:Xaa)
FT /transl_except= (pos:337..339; aa:Xaa)
FT /transl_except= (pos:340..342; aa:Xaa)
FT /transl_except= (pos:346..348; aa:Xaa)
FT /transl_except= (pos:367..369; aa:Xaa)
FT /transl_except= (pos:394..396; aa:Xaa)
FT /transl_except= (pos:397..399; aa:Xaa)
FT /transl_except= (pos:406..408; aa:Xaa)
FT /transl_except= (pos:409..411; aa:Xaa)
FT /transl_except= (pos:412..414; aa:Xaa)
FT /transl_except= (pos:421..423; aa:Xaa)
FT /transl_except= (pos:430..432; aa:Xaa)
FT /transl_except= (pos:433..435; aa:Xaa)
FT /transl_except= (pos:436..438; aa:Xaa)
FT /transl_except= (pos:442..444; aa:Xaa)
FT /transl_except= (pos:445..447; aa:Xaa)
FT /transl_except= (pos:475..477; aa:Xaa)
FT /transl_except= (pos:481..483; aa:Xaa)
FT /transl_except= (pos:487..489; aa:Xaa)
FT /transl_except= (pos:490..492; aa:Xaa)
FT /transl_except= (pos:493..495; aa:Xaa)
FT /transl_except= (pos:496..498; aa:Xaa)
FT /transl_except= (pos:502..504; aa:Xaa)
FT /transl_except= (pos:514..516; aa:Xaa)
FT /transl_except= (pos:529..531; aa:Xaa)
FT /transl_except= (pos:535..537; aa:Xaa)
FT /transl_except= (pos:562..564; aa:Xaa)
FT /transl_except= (pos:589..591; aa:Xaa)
FT /transl_except= (pos:595..597; aa:Xaa)
FT /transl_except= (pos:598..600; aa:Xaa)
FT /transl_except= (pos:604..606; aa:Xaa)
FT /transl_except= (pos:619..621; aa:Xaa)
FT /transl_except= (pos:634..636; aa:Xaa)
FT /transl_except= (pos:655..657; aa:Xaa)
FT /transl_except= (pos:691..693; aa:Xaa)
FT /transl_except= (pos:694..696; aa:Xaa)
FT /transl_except= (pos:700..702; aa:Xaa)
FT /transl_except= (pos:703..705; aa:Xaa)
FT /transl_except= (pos:706..708; aa:Xaa)
FT /transl_except= (pos:712..714; aa:Xaa)
FT /transl_except= (pos:715..717; aa:Xaa)
FT /transl_except= (pos:724..726; aa:Xaa)
FT /transl_except= (pos:727..729; aa:Xaa)
FT /transl_except= (pos:736..738; aa:Xaa)
FT /transl_except= (pos:748..750; aa:Xaa)
FT /transl_except= (pos:757..759; aa:Xaa)
FT /transl_except= (pos:760..762; aa:Xaa)
FT /transl_except= (pos:769..771; aa:Xaa)
FT /transl_except= (pos:772..774; aa:Xaa)
FT /transl_except= (pos:775..777; aa:Xaa)
FT /transl_except= (pos:778..780; aa:Xaa)
FT /transl_except= (pos:790..792; aa:Xaa)
FT /transl_except= (pos:793..795; aa:Xaa)
FT /transl_except= (pos:811..813; aa:Xaa)
FT /transl_except= (pos:820..822; aa:Xaa)
FT /transl_except= (pos:847..849; aa:Xaa)
FT /transl_except= (pos:850..852; aa:Xaa)
FT /transl_except= (pos:853..855; aa:Xaa)
FT /transl_except= (pos:856..858; aa:Xaa)
FT /transl_except= (pos:862..864; aa:Xaa)
FT /transl_except= (pos:880..882; aa:Xaa)
FT /transl_except= (pos:883..885; aa:Xaa)
FT /transl_except= (pos:892..894; aa:Xaa)
FT /transl_except= (pos:922..924; aa:Xaa)
FT /transl_except= (pos:928..930; aa:Xaa)
FT /transl_except= (pos:943..945; aa:Xaa)
FT /transl_except= (pos:952..954; aa:Xaa)
FT /transl_except= (pos:955..957; aa:Xaa)
FT /transl_except= (pos:970..972; aa:Xaa)
FT
```


851 CYATKYTRAAATCMCTGCCACTTTGGGCTATYGTMTGTGCACAYTTTTCT 900
301 TyAsnTTPThrPheTyThrLeuLeuThrLeuLeuProThrTyMetLy 317
|||||
901 TACAACTGGACTTTTATACATTTTGTGACMTATTCCTACTATATGAA 950
|||||
317 sGluValLeuArgPheAsnIleGlnGlnAsnGlyPheLeuSerAlaValP 334
|||||
951 GGARRCTCAAGTTCAATRTTCAAGACAATGGGTTTTTATCTCAKATSC 1000
|||||
334 roTyLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaasp 350
|||||
1001 CTTATTTAGGTTSTGGTTATGTATGATCCTGCTCKGGTCAAGCTGCTGAC 1050
|||||
351 AsnLeuArgAlaArgTTPAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1051 AATTAAAGGCAARATGGAATTTTCAACTTTRTGKGTTCMGAGARTTTT 1100
|||||
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1101 TAGCCTTATAGGATGATTGGACCTGCRRTATTCTGCTGTCGCGWGMT 1150
|||||
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1151 TYATWGGCTGTGATATATTCYTTGGCYGTTCCTTAACATATATCAACA 1200
|||||
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1201 ACMCTGGAGGCTTTTGCTCTCTGGATTTAGCATCAACCATCTGGAYAT 1250
|||||
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlat 434
|||||
1251 TGCTCTCTCGTATGCTGTATYCTCTGGGCATCACAAATACMTTGGCA 1300
|||||
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1301 CTATTCCKGSAATGRTTGGGCCCTCATCTGCYARAAGTCTKACCCCTGAK 1350
|||||
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1351 AACACTRTTGGAGATGCGCAACVGTGTTCTRYATYGTCTGCTGTATYAA 1400
|||||
467 nValPheGlyAlaIlePhePheThrLeuPheAlaGlyGluValGlnA 484
|||||
1401 TGTWTTTGGTGCCATTTCCTTYACACTATTGCGCAAGGTCAAGTRCAA 1450
|||||
484 snTrpAlaIleSerAspHisGlnGlyHisArg 494
|||||
1451 ACTGGCYMTCAATCATCACCAGWGCACACAGA 1482
|||||

seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.AAZ50881

seq_documentation_block:

ID AAZ50881 standard; cDNA; 1485 BP.

XX AC AAZ50881;

XX DT 31-MAY-2000 (first entry)

XX DE Human/Sheep consensus GBS toxin receptor cDNA sequence-2.

XX KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.

XX OS Homo sapiens.

XX OS Ovis sp.

XX FH Key Location/Qualifiers

FT mat_peptide 1..1485

/tag= a
/product= "Human/sheep consensus GBS toxin receptor
protein sequence"
/transl_except= (pos:4..6; aa:Xaa)
/transl_except= (pos:7..9; aa:Xaa)
/transl_except= (pos:16..18; aa:Xaa)
/transl_except= (pos:22..24; aa:Xaa)
/transl_except= (pos:28..30; aa:Xaa)
/transl_except= (pos:31..33; aa:Xaa)
/transl_except= (pos:34..36; aa:Xaa)
/transl_except= (pos:46..48; aa:Xaa)
/transl_except= (pos:49..51; aa:Xaa)
/transl_except= (pos:58..60; aa:Xaa)
/transl_except= (pos:61..63; aa:Xaa)
/transl_except= (pos:64..66; aa:Xaa)
/transl_except= (pos:67..69; aa:Xaa)
/transl_except= (pos:70..72; aa:Xaa)
/transl_except= (pos:73..75; aa:Xaa)
/transl_except= (pos:79..81; aa:Xaa)
/transl_except= (pos:85..87; aa:Xaa)
/transl_except= (pos:91..93; aa:Xaa)
/transl_except= (pos:100..102; aa:Xaa)
/transl_except= (pos:124..126; aa:Xaa)
/transl_except= (pos:130..132; aa:Xaa)
/transl_except= (pos:136..138; aa:Xaa)
/transl_except= (pos:154..156; aa:Xaa)
/transl_except= (pos:157..159; aa:Xaa)
/transl_except= (pos:163..165; aa:Xaa)
/transl_except= (pos:169..171; aa:Xaa)
/transl_except= (pos:181..183; aa:Xaa)
/transl_except= (pos:187..189; aa:Xaa)
/transl_except= (pos:190..192; aa:Xaa)
/transl_except= (pos:196..198; aa:Xaa)
/transl_except= (pos:202..204; aa:Xaa)
/transl_except= (pos:211..213; aa:Xaa)
/transl_except= (pos:220..222; aa:Xaa)
/transl_except= (pos:223..225; aa:Xaa)
/transl_except= (pos:235..237; aa:Xaa)
/transl_except= (pos:241..243; aa:Xaa)
/transl_except= (pos:244..246; aa:Xaa)
/transl_except= (pos:250..252; aa:Xaa)
/transl_except= (pos:277..279; aa:Xaa)
/transl_except= (pos:280..282; aa:Xaa)
/transl_except= (pos:283..285; aa:Xaa)
/transl_except= (pos:295..297; aa:Xaa)
/transl_except= (pos:304..306; aa:Xaa)
/transl_except= (pos:337..339; aa:Xaa)
/transl_except= (pos:340..342; aa:Xaa)
/transl_except= (pos:346..348; aa:Xaa)
/transl_except= (pos:367..369; aa:Xaa)
/transl_except= (pos:394..396; aa:Xaa)
/transl_except= (pos:397..399; aa:Xaa)
/transl_except= (pos:406..408; aa:Xaa)
/transl_except= (pos:409..411; aa:Xaa)
/transl_except= (pos:412..414; aa:Xaa)
/transl_except= (pos:421..423; aa:Xaa)
/transl_except= (pos:430..432; aa:Xaa)
/transl_except= (pos:433..435; aa:Xaa)
/transl_except= (pos:436..438; aa:Xaa)
/transl_except= (pos:442..444; aa:Xaa)
/transl_except= (pos:445..447; aa:Xaa)
/transl_except= (pos:463..465; aa:Xaa)
/transl_except= (pos:475..477; aa:Xaa)
/transl_except= (pos:481..483; aa:Xaa)
/transl_except= (pos:487..489; aa:Xaa)
/transl_except= (pos:490..492; aa:Xaa)
/transl_except= (pos:493..495; aa:Xaa)
/transl_except= (pos:496..498; aa:Xaa)
/transl_except= (pos:502..504; aa:Xaa)
/transl_except= (pos:514..516; aa:Xaa)
/transl_except= (pos:529..531; aa:Xaa)
/transl_except= (pos:535..537; aa:Xaa)
/transl_except= (pos:562..564; aa:Xaa)

FT /transl_except= (pos:589..591; aa:Xaa)
FT /transl_except= (pos:595..597; aa:Xaa)
FT /transl_except= (pos:598..600; aa:Xaa)
FT /transl_except= (pos:604..606; aa:Xaa)
FT /transl_except= (pos:619..621; aa:Xaa)
FT /transl_except= (pos:634..636; aa:Xaa)
FT /transl_except= (pos:655..657; aa:Xaa)
FT /transl_except= (pos:691..693; aa:Xaa)
FT /transl_except= (pos:694..696; aa:Xaa)
FT /transl_except= (pos:703..705; aa:Xaa)
FT /transl_except= (pos:706..708; aa:Xaa)
FT /transl_except= (pos:712..714; aa:Xaa)
FT /transl_except= (pos:715..717; aa:Xaa)
FT /transl_except= (pos:724..726; aa:Xaa)
FT /transl_except= (pos:727..729; aa:Xaa)
FT /transl_except= (pos:736..738; aa:Xaa)
FT /transl_except= (pos:748..750; aa:Xaa)
FT /transl_except= (pos:757..759; aa:Xaa)
FT /transl_except= (pos:760..762; aa:Xaa)
FT /transl_except= (pos:769..771; aa:Xaa)
FT /transl_except= (pos:772..774; aa:Xaa)
FT /transl_except= (pos:775..777; aa:Xaa)
FT /transl_except= (pos:778..780; aa:Xaa)
FT /transl_except= (pos:790..792; aa:Xaa)
FT /transl_except= (pos:793..795; aa:Xaa)
FT /transl_except= (pos:811..813; aa:Xaa)
FT /transl_except= (pos:820..822; aa:Xaa)
FT /transl_except= (pos:847..849; aa:Xaa)
FT /transl_except= (pos:850..852; aa:Xaa)
FT /transl_except= (pos:853..855; aa:Xaa)
FT /transl_except= (pos:856..858; aa:Xaa)
FT /transl_except= (pos:862..864; aa:Xaa)
FT /transl_except= (pos:880..882; aa:Xaa)
FT /transl_except= (pos:883..885; aa:Xaa)
FT /transl_except= (pos:892..894; aa:Xaa)
FT /transl_except= (pos:922..924; aa:Xaa)
FT /transl_except= (pos:928..930; aa:Xaa)
FT /transl_except= (pos:943..945; aa:Xaa)
FT /transl_except= (pos:952..954; aa:Xaa)
FT /transl_except= (pos:955..957; aa:Xaa)
FT /transl_except= (pos:970..972; aa:Xaa)
FT /transl_except= (pos:994..996; aa:Xaa)
FT /transl_except= (pos:997..999; aa:Xaa)
FT /transl_except= (pos:1009..1011; aa:Xaa)
FT /transl_except= (pos:1012..1014; aa:Xaa)
FT /transl_except= (pos:1033..1035; aa:Xaa)
FT /transl_except= (pos:1063..1065; aa:Xaa)
FT /transl_except= (pos:1081..1083; aa:Xaa)
FT /transl_except= (pos:1084..1086; aa:Xaa)
FT /transl_except= (pos:1090..1092; aa:Xaa)
FT /transl_except= (pos:1096..1098; aa:Xaa)
FT /transl_except= (pos:1111..1113; aa:Xaa)
FT /transl_except= (pos:1126..1128; aa:Xaa)
FT /transl_except= (pos:1129..1131; aa:Xaa)
FT /transl_except= (pos:1138..1140; aa:Xaa)
FT /transl_except= (pos:1141..1143; aa:Xaa)
FT /transl_except= (pos:1144..1146; aa:Xaa)
FT /transl_except= (pos:1147..1149; aa:Xaa)
FT /transl_except= (pos:1150..1152; aa:Xaa)
FT /transl_except= (pos:1153..1155; aa:Xaa)
FT /transl_except= (pos:1168..1170; aa:Xaa)
FT /transl_except= (pos:1174..1176; aa:Xaa)
FT /transl_except= (pos:1180..1182; aa:Xaa)
FT /transl_except= (pos:1189..1191; aa:Xaa)
FT /transl_except= (pos:1201..1203; aa:Xaa)
FT /transl_except= (pos:1246..1248; aa:Xaa)
FT /transl_except= (pos:1270..1272; aa:Xaa)
FT /transl_except= (pos:1291..1293; aa:Xaa)
FT /transl_except= (pos:1306..1308; aa:Xaa)
FT /transl_except= (pos:1315..1317; aa:Xaa)
FT /transl_except= (pos:1324..1326; aa:Xaa)
FT /transl_except= (pos:1330..1332; aa:Xaa)

FT /transl_except= (pos:1333..1335; aa:Xaa)
FT /transl_except= (pos:1339..1341; aa:Xaa)
FT /transl_except= (pos:1348..1350; aa:Xaa)
FT /transl_except= (pos:1357..1359; aa:Xaa)
FT /transl_except= (pos:1372..1374; aa:Xaa)
FT /transl_except= (pos:1375..1377; aa:Xaa)
FT /transl_except= (pos:1381..1383; aa:Xaa)
FT /transl_except= (pos:1384..1386; aa:Xaa)
FT /transl_except= (pos:1396..1398; aa:Xaa)
FT /transl_except= (pos:1402..1404; aa:Xaa)
FT /transl_except= (pos:1420..1422; aa:Xaa)
FT /transl_except= (pos:1444..1446; aa:Xaa)
FT /transl_except= (pos:1456..1458; aa:Xaa)
FT /transl_except= (pos:1459..1461; aa:Xaa)
FT /transl_except= (pos:1462..1464; aa:Xaa)
FT /transl_except= (pos:1471..1473; aa:Xaa)
FT /transl_except= (pos:1483..1485; aa:Xaa)
FT /note= "Xaa = unknown"
XX
PN WO200005375-A1.
XX
XX 03-FEB-2000.
PD
XX 22-JUL-1999; 99WO-US16676.
PF
XX 22-JUL-1998; 98US-0093843.
PR
XX (UYVA-) UNIV VANDERBILT.
PA
XX Hellerqvist CG, Fu C;
PI
XX WPI; 2000-205377/18.
DR P-PSDB; AAY45090.
XX
alignment_scores:
Quality: 1971.00 Length: 494
Ratio: 4.879 Gaps: 0
Percent Similarity: 81.781 Percent Identity: 78.138
alignment_block:
US-09-776-865-4 x AAZ50881 ..
Align seg 1/1 to: AAZ50881 from: 1 to: 1485
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlyse 17
1 ATGANGTCNCCGGTTNNGACNTNGCCNGANGCANGCGGAGGAGNGCNC 50
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProv 34
51 GGACCGCACNCCNCTNCTNCGNGCCGCCCGCGGCGGANGCGCTCCAG 100
34 alCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
101 TNGTCTGCTCTGCTGCTACACNTAGCANTTTTGNCTTTTGTGTTTC 150
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValaspme 67
151 TTNCTNTNTATNCTACNGTGGAATCTGAGNGTTCGNNTAGTGGANAT 200
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysa 84
201 GGTNGATTCAANACAACCTNNNAAGATAATACAACTCCNANGNGTGN 250
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
251 CAGAGCATTTCTGCTCCATAAAAGTTCTCANAANCAAAACGGTAANA 300
101 TyrArgTyrAspAlaGluThrGlnGlyTyrIleLeuGlySerPhePhe 117
301 TACCNTTGGGATCAGAAACTCAAGGATGGATTCTCGGNTCNTTTTNTA 350

117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
351 TGGCTACATCATCACANATCTCGGAGGATATGTTGCCAGCAANAAG 400
134 lyGlyLysLeuLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
1401 GGGGGAANNTGNTCTAGGATTTGGGATCTNTGNACNGCTNTCNTCACC 450
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
451 CTGTTCACTCCCTNGCTGCAGATTTGGGAGTGGAGCCTNTTGNAGN 500
167 uArgAlaLeuGlyLeuGlyGlyValThrTyrProAlaMethHisA 184
501 CAGNGCACTAGAAGGCTAGGAGAGGCTGNACATNTCCAGCCATGCATG 550
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuSer 200
551 CCATGTGGCTCTCTNGGGCTCCCTCTTGAAGAACAACNTTCTTGNAGN 600
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
601 ATTCNTATGCGAGGACACANCTTGGGACAGATNTTCTCTCTCTCTTC 650
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
651 TGGANTAATTGCTACTATATGAATGGACTTATGTCTCTCTANTTTTG 700
234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
701 GNANTNTTGGAAATNTTGGTTNTTNTTNGGATCTGNTTAGTAGTAGAN 750
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
751 ACACANAANAACACAGANAATNCCNNNTATGAAGAAGGANTANATCT 800
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
801 TTCATCATTAANAATACAGCTNTCTTACAGAGTCAAGTCCCGTGGTAC 850
284 roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
851 CNATPNTNAAATCCTGCCACTTGGGCTATNGTNGTGTGCACANTTTTCT 900
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
901 TACAACCTGGACCTTTTATACTTTTGGACNTTATGCTCTACTTANATGAA 950
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
951 GGANNTCCTAAGGTTCAATNTTCAAGAGATGGGTTTTATCNCANTNC 1000
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
1001 CTTATTTAGGNTNTGGTTATGTATGATCTCTGTCNGCTCAAGCTGCTGAC 1050
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgValPh 367
1051 AATTTAAGGCAANATGGAAATTTTCAACCTNTNGNTTTCGNAGANTTTT 1100
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
1101 TAGCTTTATAGGNATGATTGGACCTGCNNATTTCTCTGGTNGCNGCNGNT 1150
384 helIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
1151 T NATNGGCTGTGATTTATCTNTGGCNGTTGCTTCTTAACNATATCAACA 1200
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuaspII 417
1201 ACNCTGGAGGCTTTTCTCTCTGGATTTAGCATCAACCATCTGGANAT 1250
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlat 434

1251		TGCTCCTCGTATGCTGGTATNCTCTGGGCATCAAAATACNTTTGCCA	1300
434	hrtleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu	450	
1301	CTATTCGCGAATGTTGGGCCNTCAATGCGNANAAGTCTNACCCCTGAN	1350	
451	AsnThrIleGlyGluTTPGlnThrValPheCysIleAlaAlaIleAs	467	
1351	AACACTNTTGGAGATGCGAAACNGTNTCTNNATNGCTGCTGCTATNAA	1400	
467	nValPheGlyAlaIlePhePheThrLeuPheAlaIleGlyGluValGlnA	484	
1401	TGNTTTTGGTGCCATTTCCTNACACTATTGCGCAAGGTGAAGTNCARA	1450	
484	snTrpAlaIleSerAspHisGlnGlyHisArg	494	
1451	ACTGGGCNNTCANTGATCACCANGACACAGA	1482	
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT: A3A			
seq_documentation_block:			
ID	AAH79234 standard; cDNA; 2670 BP.		
XX			
AC	AAH79234;		
XX			
DT	20-NOV-2001 (first entry)		
XX			
DE	Human sodium dependent phosphate co-transfer protein 35 cDNA.		
XX			
KW	Human; sodium dependent phosphate co-transfer protein 35;		
KW	hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;		
KW	nephritis; gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	CNT1298882-A.		
XX			
PD	13-JUN-2001.		
XX			
PF	06-DEC-1999; 99CN-0124217.		
XX			
PR	06-DEC-1999; 99CN-0124217.		
XX			
PA	(SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.		
XX			
PI	Mao Y, Xie Y;		
XX			
DR	WPI; 2001-503367/56.		
DR	P-PSDB; AAG65238.		
XX			
PT	Human Na-dependent phosphate cotransporter 35 and its coding		
XX			
PS	Claim 6; Page 19-20(Disclosure); 28pp; Chinese.		
XX			
CC	The present invention provides the protein and coding sequence		
CC	sodium dependent phosphate co-transfer protein 35. The sequence		
CC	used in the treatment of hypophosphaturia, hypercalcaemia,		
CC	hypophosphataemic rickets and nephritis. The present sequence		
CC	coding sequence of the invention.		
XX			
SQ	Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;		

```

alignment_scores:
  Quality: 1578.00      Length: 321
  Ratio: 5.042          Gaps: 0
  Percent Similarity: 97.508  Percent Identity: 89.097

alignment_block:
  US-09-776-865-4 x AAH79234
  ..

Align seg 1/1 to: AAH79234 from: 1 to: 2670

```


175 GluGlyValThrTyrProAlaMetHisAlaMetTyrSerSerTyrAlaPr 191
|||||
3 GAGGGTGTACATTCCAGCCATGATGCCATGGTCTCTTGGGCTCC 52

191 oProLeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnL 208
|||||
53 CCCTCTTGAAGAAGCAAACTCTTAGCATTTTCATATGCAGAGCACAGC 102

208 euGlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMet 224
|||||
103 TTGGGACAGTAATTTCTTCCCTTCTTGGAAATAATTTGCTACTATATG 152

225 AsnTrpThrTyrValPheTyrPheGlyIleValGlyIleIleIleTrpPh 241
|||||
153 AATTGGACTTATGCTCTACTTTTGGTACTATTGGAATAATTTGGTT 202

241 eIleLeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrI 258
|||||
203 TCTTTTGTGGATCTGGTTAGTTAGTACACACACACAAAAACAAGAGAA 252

258 leThrProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnLeu 274
|||||
253 TTCCCATTAAGAAGGAATACATTTCTTCAATTAAGAATACAGCTT 302

275 SerSerGlnLysSerValProTyrIleProMetLeuLysSerLeuProLe 291
|||||
303 TCTTCACAGAAGTCAGTCGGTGGTACCATTTTAAATCCCTGCCACT 352

291 uTrpAlaIleValAlaHisPheSerTyrAsnTrpThrPheTyrThrL 308
|||||
353 TTGGGCTATCGTAGTGCACACTTTTCTTACAACTGGACTTTTATACCT 402

308 euLeuThrLeuLeuProThrTyrMetLysGluValLeuArgPheAsnIle 324
|||||
403 TATTGACATTTATGCTTACTTATGAAGGAGATCCTTAAGTTCAATGTT 452

325 GlnLysAsnGlyPheLeuSerAlaValProTyrLeuGlyCysTrpLeuCy 341
|||||
453 CAAGAGAATGGGTTTTATCTTCATTCGCCCTATTATTAGGCTCTTGGTTATG 502

341 sMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaArTrpAsnP 358
|||||
503 TATGATCCTGTCTGGTCAAGCTCTGCACAAATTTAAGGGCAAAATGGAAT 552

358 heSerThrLeuTrpValArgValPheSerLeuIleGlyMetIleGly 374
|||||
553 TTTCACTTTATGTGTTCGCAGAAATTTTAGCCCTTATAGGAATGATTGGA 602

375 ProAlaIlePheLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLe 391
|||||
603 CCTGCAGTATCTCGTAGCTGGCTTCATTTGGCTGTGATTATCTTT 652

391 uAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerS 408
|||||
653 GGCCGTTGCTTTCTTAACATATATCAACACACTGGGAGGCTTTTGCTCTT 702

408 erGlyPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIle 424
|||||
703 CTGGATTTAGCATCAACATCTGGAATTTGCTCTTCGTATGCTGGTATC 752

425 LeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetIleGlyPr 441
|||||
753 CTCCTGGGCATCAATAATCATTTGCCACTATTCAGGAATGTTGGGCC 802

441 oIleIleAlaArgSerLeuThrProGluAsnThrIleGlyGluTrpGlnT 458
|||||
803 CGTCATTGTAAAGTCTGACCCCTGATAACACTGTTGGAGAATGCGAAA 852

458 hrValPheCysIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePhe 474
|||||
853 CCGTGTCTTATATGTCTGTGCTATTAAATGTTTTTGGTGCCATTTCTTT 902

475 ThrLeuPheAlaLysGlyGluValGlnAsnTrpAlaIleSerAspHisG1 491
|||||
903 ACACATTCCCAAGGTGAAGTACAAACTGGCTCTCAATCATCACCA 952

491 nGlyHisArgAsn 495
:|||||
953 TGGACACAGACAC 965

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI59901

seq_documentation_block:
ID AAI59901 standard; cDNA; 929 BP.
XX
AC AAI59901;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3890.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
(HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
WPI: 2001-442253/47.
DR P-PSDB; AAM40745.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3890; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX

SQ Sequence 929 BP; 214 A; 224 C; 223 G; 268 T; 0 other;

alignment_scores:

Quality: 1217.00 Length: 270
Ratio: 4.810 Gaps: 0
Percent Similarity: 93.704 Percent Identity: 82.963

alignment_block:

US-09-776-865-4 x AAI59901

Align seg 1/1 to: AAI59901 from: 1 to: 929

```
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlyLeu 17
119 ATGAGGTCCTCGGTCGACAGCTGCGCCGACGATGCGGAGGAGACAC 168
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaPro 34
169 GGACCGCAGCGCTCTTCTACGGCGCGCCACGCGCGGAGCGCTCCAG 218
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
171 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 268
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
269 TTCATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 318
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCys 84
319 GGTAGATTCAATCAACTTTAGAGATAATAGAACTTCCAAAGCGGTGTC 368
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLys 100
369 CAGACGATTTCTGCTCCCAATAAGTTTCATCATATCAAGCGGTGTAAG 418
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
419 TACCAATGGGATGAGAACTCAGAGTGGATTCCTCGTTCCTTTTATA 468
117 rGlyTyrIleIlePheGlnIleProGlyGlyTyrValAlaSerArgSerG 134
469 TGGCTACATCATCACAGATTCCTGGGAGATATGTTGCCAGCAAAATAG 518
134 lGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
519 GGGGAAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGCTCCTCAC 568
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
569 CTGTTCACTCCCATGCTGAGATTTAGGAGTTGGACCACTCAATGTA 618
167 uArgAlaLeuGluGlyLeuGlyGlyValThrTyrProAlaMethisA 184
619 CAGAGCACTAGAAGGACTAGGAGGGTGTACATTTCCAGCCATGATG 668
184 laMetTrpSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
669 CCATGTGCTCTTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTAGC 718
201 lIleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
719 ATTTCAATGACAGAGCAGCTGCGGACAGTAATTTCTTCTTCTCTTC 768
217 rGlyValIleCysTyrTrpMetAsnTrpThrTyrValPheTyrPhePheG 234
769 TGAATAAATTTGCTACTATATGAATGAGCTTATGCTTCTTCTTCTTTT 818
234 lYlleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
819 GTACTATGGAATATTTTGTGTTCTTTTGTGGATCTGTTAGTAGTGAC 868
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK93901

seq_documentation_block:

ID AAK93901 standard; cDNA; 853 BP.

XX

AC AAK93901;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human cDNA clone representative sequence, SEQ ID NO: 2361.

XX

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX

OS Homo sapiens.

XX

PN EP1130094-A2.

XX

XX 05-SEP-2001.

XX

PF 07-JUL-2000; 2000EP-0114089.

XX

PR 08-JUL-1999; 99JP-0194486.

PR

PR 11-JAN-2000; 2000JP-0118774.

PR

PR 02-MAY-2000; 2000JP-0183765.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI; 2001-524255/58.

XX

PT 830 Primers useful for synthesizing full length cDNA clones and their

PT use in genetic manipulation -

XX

PS Example 11; SEQ ID NO 2361; 1380pp + sequence listing; English.

XX

CC The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence was used as the CC representative sequence from a human clone which was used in CC homology searches to identify the clone.

CC Note: the sequence data for this patent did not form part of the printed CC specification, but was obtained in CD-ROM format directly from EPO.

XX

SQ Sequence 853 BP; 198 A; 200 C; 213 G; 239 T; 3 other;

alignment_scores:

Quality: 1104.00 Length: 248
Ratio: 4.779 Gaps: 0
Percent Similarity: 93.145 Percent Identity: 82.258

alignment_block:

US-09-776-865-4 x AAK93901

Align seg 1/1 to: AAK93901 from: 1 to: 853


```
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheThy 117
    |||:::|||||
383 TACCAATGGGATGCAGAAACTCAAGATGATTCGGTTCCTTTTAA 432
    |||:::|||||
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerg 134
    |||:::|||||
433 TGGCTACATCATCACAGATTCCTGGAGGATATGTTGCCAGCAAAATAG 482
    |||:::|||||
134 lYGLyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
    |||:::|||||
483 GGGGAAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGCCACACC 532
    |||:::|||||
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
    |||:::|||||
533 CTGTTCACTCCCATGCTGCAGATTTAGGAGTTGGACCACCTATTGTACT 582
    |||:::|||||
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrValValSerProLeuSe 184
    |||:::|||||
583 CAGAGCACTAGAAGGACTANGAGAGGGTGTACATTTCCAGCCATGCATG 632
    |||:::|||||
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
    |||:::|||||
633 CCATGTGCTCTCTTGGGCTCCCCCTCTTGAAGAAGCAAACTCTTAGC 682
    |||:::|||||
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
    |||:::|||||
683 ATTTCATATCGAGGACACANCTTGGGACAGTAATTTCTTCTTCTTCTT 732
    |||:::|||||
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
    |||:::|||||
733 TGAATAAT.TGCTACTATATGAATGGGACTATGCTTCTACTTTTGTG 781
    |||:::|||||
234 lYleValGlyIleIleTrpPheIleLeuTrp 244
    |||:::|||||
782 GGACTATGGAATATTGGGCTC...FTTTGG 810
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL03769

seq_documentation_block:

ID ABL03769 standard; cDNA; 1939 BP.

AC ABL03769;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5789.

DE Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical; gene; ss.

KW Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB59666.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX

PS Claim 1; SEQ ID NO 5789; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1939 BP; 410 A; 558 C; 552 G; 419 T; 0 other;

alignment_scores:

Quality: 991.50 Length: 517

Ratio: 2.857 Gaps: 7

Percent Similarity: 67.118 Percent Identity: 39.458

alignment_block:

US-09-776-865-4 x ABL03769 ..

Align seg 1/1 to: ABL03769 from: 1 to: 1939

11 SerAspGlyGluGlySerAspArg.....Th 20

353 TCCGACGACGACGCGGAGGAGGCGTTTGTCCGGCGAAACG 402

20 rProLeuLeuGlnArgAlaProAlaGluProAlaProValCys...C 36

403 ACCGCTAATCCGCTCCAGCGGCGCTCGGAGGAGAAACCATGGCTGGCC 452

36 ysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPheVal 52

453 CAAAGACGGCTCACATTTTCGGATTTCCTGGGATTTCGGCGTG 502

53 LeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMetValas 69

503 GTCACGGATGGGGTCAATCTGCGTGGCCATTTGGCCCATGGTG... 550

69 pSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysAlaGluH 86

551AACCAACGGCAATTCGCGACAGCACTCATCGGTGATTGATACGG 596

86 isSerAlaProIleLysValLeuHisAsn..... 95

597 ACAGGTGTCCTACTACCGGCACCAATCACAATGGTAGCGATCCCAATCCG 646

96 GlnThrGlyLysLysTyrArgTrpAspAlaGluThrGlnGlyTrpIleLe 112

647 CAGAAGGAGGGGAGTTTGTGGGACGAGCCACGCGAGGATTTGGTGT 696

112 uGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyrV 129

697 CGGCAGTTTCTTCTATGGCTATGCTAACCAAGTGCCCGGGGACGGA 746

129 alAlaSerArgSerGlyLysLeuLeuLeuGlyPheGlyIlePheAla 145

747 TGGCCGAGCTGTATGTTGGGAGAGATCTACGGCTATGAGTGTGTGATC 796

146 ThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPheGlyValGl 162

797 ACGGCGGTCTTTACGCTTATACCTCATCTGGCTGGCCACTGGGATCTGCC 846

162 yAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluGlyValThrT 179

847 GCTGTTGGTCTGTCGTCATCTGCGAGGAATGGCGGAGGGCGTCACCT 896

179 yProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGluArg 195

|||||

470 GlyAlaIlePhePheThrLeuPheAlaLysClyGluValGlnasnTrpAl 486
||| : : : : : ||| : : : : : ||| : : : : :
1847 GGTAACTTCATCTACCTGATCTCGCAGCGCCGAGGACGAAGCTGGT 1896

486 a 486
:
1897 G 1897

OM of: US-09-776-865-4 to: EST:* out_format : pfs

Date: Jul 15, 2002 2:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2.1/USPTO_spool/US09776865/runat_15072002_082207_27382/app_query.fasta_1.1159
-DB=EST -QPMF=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LQOPL=0.000 -LQOPEXT=0.000 -CGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTPMF=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09776865.ecgn1.1.3468
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPY -WAIT -THREADS=1

Search information block:

Query: US-09-776-865-4

Query length: 495

Database: EST:*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 2434.030000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_est1:AL550137	+ 1395.00	2364.64	1.8e-122	985	AL550137 LTI_NFL006.PL1
gb_est2:BF661062	+ 1038.00	1756.57	1.4e-88	801	BF661062 603304362E1 NIH_CGAP_M
gb_est2:BF907284	+ 953.00	1611.92	1.5e-80	754	BF907284 603063858F1 NIH_MGC_11
gb_est2:BF676817	+ 921.00	1559.22	1.3e-77	626	BF676817 602084380F1 NIH_MGC_83
gb_est2:BF697765	+ 915.00	1547.59	5.9e-77	711	BF697765 603346858F1 NCI_CGAP_M
gb_hic:BC023331	- 900.00	1513.67	4.6e-75	1531	BC023331 Mus musculus, clone I
gb_est2:BF650521	+ 879.50	1486.53	1.5e-73	736	BF650521 603386787F1 NIH_MGC_87
gb_est2:BF539146	+ 868.00	1464.76	2.4e-72	895	BF539146 602053009F1 NCI_CGAP_M
gb_est1:BF537525	+ 839.50	1419.57	8.0e-70	645	BF537525 BF537525 RIKEN full-le
gb_est2:BF124137	+ 837.00	1411.65	2.2e-69	904	BF124137 601760838F1 NCI_CGAP_M
gb_est1:BF613552	+ 821.00	1387.34	5.0e-68	682	BF613552 BF613552 RIKEN full-le
gb_est2:BF654109	+ 808.00	1362.21	1.3e-66	893	BF654109 602570265F1 NIH_MGC_77
gb_est1:AA833297	+ 803.00	1359.54	1.8e-66	518	AA833297 ud05d09.r1 Soares NMPU
gb_est2:BF187031	+ 796.00	1345.41	1.1e-65	633	BF187031 UMN07B04 Canine Brain
gb_est2:BF867611	+ 782.00	1320.77	2.5e-64	676	BF867611 601443127F1 NIH_MGC_65
gb_est2:BF869819	+ 736.00	1241.75	6.4e-60	700	BF869819 604446651F1 NIH_MGC_65
gb_est2:BF181890	+ 716.50	1209.09	4.2e-58	658	BF181890 603378047F1 NCI_CGAP_M
gb_est2:BF866603	+ 709.00	1190.72	4.5e-57	1100	BF866603 602785577F1 NCI_CGAP_M
gb_est1:BF610013	+ 698.00	1177.93	2.3e-56	630	BF610013 BF610013 RIKEN full-le
gb_est1:BF664731	+ 697.50	1177.09	2.6e-56	629	BF664731 BF664731 RIKEN full-le
gb_est2:BF179717	+ 694.50	1172.71	4.5e-56	587	BF179717 da11a1a.y1 NICHHD XGC
gb_est2:BF060500	+ 673.00	1133.73	6.7e-54	721	BF060500 BF060500 NIBB Mochii r
gb_est2:BF563945	+ 662.00	1120.53	3.6e-53	429	BF563945 UI-R-C4-akw-c-03-0-UI
gb_est1:AA285133	+ 638.00	1079.43	7.1e-51	432	AA285133 zr59d01.r1 Soares NHHM
gb_est2:BF031727	+ 624.00	1051.39	2.6e-49	632	BF031727 BF031727 NIBB Mochii r
gb_est1:AL597124	+ 576.00	971.88	6.9e-45	499	AL597124 DKEP2313H91F1 r1.313
gb_est2:BF375421	+ 538.00	908.16	2.4e-41	445	BF375421 601230485F1 NCI_CGAP_M
gb_est2:BF445533	+ 529.00	886.73	3.0e-40	847	BF445533 daebln09.y3 NICHHD XGC
gb_est2:BF499662	+ 529.00	886.29	4.0e-40	811	BF499662 AT14282.5prime AT Dros
gb_gss:CN50069V	+ 506.00	843.93	9.2e-38	1075	AL063986 Drosophila melanogast
gb_est2:BF1873013	+ 504.00	842.85	1.1e-37	866	BF1873013 603398193F1 NIH_MGC_94
gb_est2:BF123545	+ 500.50	838.10	2.0e-37	772	BF123545 601759862F1 NCI_CGAP_M
gb_est2:BF825556	+ 497.00	832.29	4.1e-37	760	BF825556 601433358F1 NIH_MGC_72
gb_est2:BF3685	+ 494.50	835.87	2.6e-37	367	BF3685 yr55601.r1 Soares fetal
gb_est2:BF1734022	+ 481.50	805.89	1.2e-35	753	BF1734022 603355577F1 NIH_MGC_94
gb_gss:CN50352N	+ 475.00	792.35	6.9e-35	943	AL229388 Tetraodon nigroviridis
gb_est2:BF1753164	- 465.50	779.95	3.4e-34	661	BF1753164 603026028F1 NIH_MGC_11
gb_est2:BF495471	+ 465.50	779.65	3.5e-34	679	BF495471 AT04410.5prime AT Dros
gb_est2:BF13524	+ 453.00	767.07	1.8e-33	301	N31254 YR53H06.r1 Soares meland
gb_est2:BF502767	+ 435.50	727.93	2.7e-31	707	BF502767 AT18580.5prime AT Dros

gb_est2:BE977644 + 433.50 725.67 3.6e-31 635 ! BE977644 bs67c07.y1 Drosophi
gb_est2:BF505525 + 431.50 719.95 7.4e-31 786 ! BF505525 AT07766.5prime AT D
gb_est1:AW531337 - 431.00 728.77 2.4e-31 321 ! AW531337 UI-R-C4-akw-c-03-0-
gb_est2:BG296175 + 430.00 717.43 1.0e-30 783 ! BG296175 602394006F1 NIH_MGC
gb_est2:BG123873 + 427.50 713.46 1.7e-30 761 ! BG123873 ES1469519 tomato sh
seq_name: gb_est1:AL550137
seq_documentation_block:
LOCUS AL550137 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.PL2 Homo sapiens cDNA clone CSODI040YK17.5
prime, mRNA sequence.
ACCESSION AL550137
VERSION AL550137.1 GI:128868913
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..985
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSODI040YK17"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fillang@life.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 236 a 224 c 224 g 300 t
ORIGIN

alignment_scores:

Quality: 1395.00 Length: 309
Ratio: 4.777 Gaps: 0
Percent Similarity: 94.498 Percent Identity: 83.819

alignment_block:

US-09-776-865-4 x AL550137
Align seg 1/1 to: AL550137 from: 1 to: 985
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlyse 17
60 ATGAGGTCCTCGGTGAGACCTGCGGCGGACGAGGAGGAGGAGCAC 109
17 rAspArgThrProLeuLeuGluAlaArgAlaProArgAlaGluProAlaProV 34
110 GGACCGGACCGCCTCTCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCAG 159
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyphe 50
160 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 209
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
210 TTCATTGCTGATGCATTACGTTGTAATCTGAGTGTGCTGCTGCTGCTGCTGCTGCTAT 259


```
352 GGGTGGATTCTCGGCTCTTTTTTTAGCGGTACATCTCACCAGATTCC 401
125 oglyGlyTyrValAlaSerArgSerGlyGlyLysLeuLeuGlyPheG 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 CGGTGGGTACATTGCCAGAGGTGCGAGGGAAGCTGCTGCTGGCCCTGG 451
142 lylePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAsp 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 GCATCTTAGGCACCTCGCTTCCACCTGTTCACCGCTGCGCGGCAGAC 501
159 PheGlyValGlyAlaLeuValAlaArgAlaLeuGlyGlyGlyGlyGly 175
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 TTAGCGGTGTGACTTCGTTGGCTTAGAGCGCTGAAGGACTGGGAGA 551
175 uGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaProp 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 GGGTGTACGCTTCCAGCTATGACGCGCATGTGCTCTTCCTGGGCTCCC 601
192 roLeuGluArgSerLysLeuLeuSerLysSerTyrAlaGlyAlaGlnLeu 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 CTCTGGAAAGCAAGCTTCTTACCAATTTCCTATGCGGAGCACAGCTT 651
209 GlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMetAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
652 GGCAGCTGATCTCACTTCTCTTTCGGAAATATATGCTACTATATGAA 701
225 nTrpThrTyrValPheTyr.PhePheGlyIleValGlyIleIleTrp.Ph 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
702 CTGGACTTACGCTCTCTATCTTT...GCTATAGTTGGAATTGCTGTTA 748
241 eLleLeuTrpIleCysLeuValSerAspThrProGlu 253
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
749 T...TTATGGATGGGATAGTACTAGTATACACCAGAA 782
```

seq_name: gb_est2:BI907284

seq_documentation_block: 754 bp mRNA linear EST 16-OCT-2001
LOCUS BI907284 603063858F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212766 5',
DEFINITION mRNA sequence.

ACCESSION BI907284
VERSION BI907284.1 GI:16170094

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 754)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLNL1534 row: d column: 15

High quality sequence stop: 750.

Location/Qualifiers

1..754

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5212766"

/clone_lib="NIH_MGC_118"

/tissue_type="Leukocyte"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV

(destroyed); RNA source leukocytes from anonymous pool of

non-activated adult donors. Library is oligo-dT primed

and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

BASE COUNT 171 a 182 c 194 g 207 t
ORIGIN

alignment_scores:

Quality: 953.00 Length: 227

Ratio: 4.453 Gaps: 1

Percent Similarity: 94.273 Percent Identity: 83.260

alignment_block:

US-09-776-865-4 x BI907284 ..

Align seg 1/1 to: BI907284 from: 1 to: 754

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlyse 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

78 ATGAGGTCTCGGTTTCGAGACCTGGCCCGGAACGATGCGAGGAGACAC 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
: ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

128 GGACCG.ACGCCTCTTACCGGGCGGCCACGGCGCGAAGCCGCTCCAG 176
: ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

177 TGTGCTGCTCTGCTCGTTACAACCTTAGCAATTTTGGCCTTTTGGTTTC 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

227 TTCATTGCTGATGCATTACGTGTAATCTGAGTTTCGCTTAGTGATAT 276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

277 GGTAGATTCAATACAACTTTAGAAGATAATAGAACTTCCAAGCGGTGTC 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

327 CAGAGCATCTGCTCCCAAGTTTCATATAATCAACGGGTAAAG 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

377 TACCAATGGGATGCAGAACTCAAGGATGGATTCTCGTTCCTTTTAA 426
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

427 TGGCTACATCATCACAGATTCTCGGAGGATATGTTGCCAGCAAAATAG 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

477 GGGGAAATGCTGCTAGGATGGGATTCCTTGGCACTGCTGCTCACC 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

527 CTGTTCACTCCCAATGCTGCAGATTTAGGAGTTGGACCACTCAATTGTA 576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

167 uArgAlaLeuGlyLeuGlyGluGlyValThrTyrProAlaMetHisA 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

577 CAGAGCACTAGAGGACTAGGAGGGGTGTACATTTCCAGCCATGCATG 626
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

184 laMetTrpSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

627 CCATGTGGTCTTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTAGC 676
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValVal.SerLeuProLys 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

677 ATTTTCATATGCAGGAGCACAGCTTGGACAGTAATTTTCTCTCTCTTT 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

217 erGlyValIleCysTyrTyrMetAsnTrp 226


```
/dev_stage="5 months"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: Sali;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT      147 a 191 c 199 g 174 t
ORIGIN

alignment_scores:
  Quality: 915.00      Length: 216
  Ratio: 4.530         Gaps: 0
  Percent Similarity: 93.519  Percent Identity: 82.407

alignment_block:
US-09-776-865-4 x BI697765 ..

Align seg 1/1 to: BI697765 from: 1 to: 711

9 AlaProSerAspGlyGluGluGlySerAspArgThrProLeuLeuGlnAr 25
||| :||| |||||:||||| ||||| ||||| |||||
64 GCGGAACGACGATGAGGAGCTGGACAGCACCCGCTCTGCGCGG 113
||| :||| |||||:||||| ||||| ||||| |||||
25 gAlaProArgAlaGluProAlaProValCysSerAlaArgTyrAsnL 42
||| :||| |||||:||||| ||||| ||||| |||||
114 CGCCCGGACAGACCGAAGCGCTCCAGTGTGCTCTGCTCGGTACAACT 163
42 euAlaPheLeuSerPhePheGlyPhePheValLeuTyrSerLeuArgVal 58
|||||:|||||:||||| ||||| ||||| ||||| |||||
164 TAGCGATTTCGGCTTCGTGGTTCCTCTCTATGCTTACGGGTG 213
59 AsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThrAlaLy 75
|||||:|||||:||||| ||||| ||||| ||||| |||||
214 AACCTGAGTGTGGTAGTGACATGCTAGATTCAATACAACTCTCAC 263
75 sAspAsnArgThrSerTyrGluCysAlaGluHisSerAlaProIleLysV 92
|||||:|||||:||||| ||||| ||||| ||||| |||||
264 TGATTAATAGAACGCTCTAAGAGAGTGTGGGAACATTCGCCGCCATAAAG 313
92 alLeuHisAsnGlnThrGlyLysLysTyrArgTyrAspAlaGluThrGln 108
||| |||||:|||||:||||| ||||| ||||| ||||| |||||
314 TTCACCACATCACACAGTAAAAGTACAAGTGGATGCGAAGAACTCAA 363
109 GlyTrpIleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIlePr 125
|||||:|||||:||||| ||||| ||||| ||||| |||||
364 GGGTGGATTCTCGGCTC.TTTTTTACGGCTACATCGTACCACCATTC 412
125 oGlyGlyTyrValAlaSerArgSerGlyGlyLysLeuLeuGlyPheG 142
|||||:|||||:||||| ||||| ||||| ||||| |||||
413 CGGTGGGTACATTGCCAGCAGGTGCGGAGGAAGCTGCTGGGCGCTGG 462
142 lyIlePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAsp 158
|||||:|||||:||||| ||||| ||||| ||||| |||||
463 GCATCTTAGGCACCTCGGCTCTCCCTGTTCCACCGCTGGCGCGACAG 512
159 PheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyG 175
:|||||:|||||:||||| ||||| ||||| ||||| |||||
513 TTAGCGGTGGTACTCTCTGTGTCTAGACGCGCTGGAGGAGACTGGAGA 562
175 uGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaProp 192
|||||:|||||:||||| ||||| ||||| ||||| |||||
563 GGGTGTACGTTTCCAGCTATGCAGCCCATGCTGCTTCCTTCCTGGGCTCCC 612
192 roLeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeu 208
|||||:|||||:||||| ||||| ||||| ||||| |||||
613 CTCTGGAAGAAGCAAGCTCTTACCATTTCTATGCGGGAGCACAGCTT 662
209 GlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMet 224
|||||:|||||:||||| ||||| ||||| ||||| |||||
663 GGGACAGTATCTACATTCCTCTTCCGGAATAATATGCTACTATTGG 710

seq_name: gb_htc:BC023331
```

```
seq_documentation_block:
LOCUS      BC023331                1531 bp      mRNA      linear      HTC 05-FEB-2002
DEFINITION Mus musculus, clone IMAGE:4500938, mrna.
ACCESSION  BC023331
VERSION    BC023331.1  GI:18497348
KEYWORDS   HTC.
SOURCE      house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 1531)
AUTHORS    Strausberg,R.
TITLE      Direct Submission
JOURNAL    Submitted (05-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAC Plate: 29 Row: f Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
            This clone has the following problem: no polyA-tail.
FEATURES             Location/Qualifiers
     source            1..1531
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4500938"
                     /tissue_type="Eye, retina, mouse strain C57BL/6"
                     /clone_lib="NIH_MGC_94"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      371 a 458 c 434 g 268 t
ORIGIN

alignment_scores:
  Quality: 900.00      Length: 459
  Ratio: 2.719         Gaps: 4
  Percent Similarity: 72.113  Percent Identity: 38.998

alignment_block:
US-09-776-865-4 x BC023331/rev ..

Align seg 1/1 to reverse of: BC023331 from: 1 to: 1531

30 GluProAlaProValCysSerAla.....ArgTyrAs 41
:||||| ||| |||:|||||
1330 GACCGCGCTGTGGTGGACTGCACCTGTGGCTTCCTCGCTCGCTACAT 1281
41 nLeuAlaPheLeuSerPhePheGlyPhePheValLeuTyrSerLeuArgV 58
:|||||:|||||:||||| ||||| ||||| ||||| |||||
1280 CATGCCCATCATGAGCGGTCTGGGTTTCTGTATCATGCTTTGGCATCCGCT 1231
58 alAsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThrAla 74
|||||:|||||:||||| ||||| ||||| ||||| |||||
```

```

1230 GCAACTGGCGTGGCCATCGTGTCCATGGTCAACAACAGCACAACCCAC 1181
75  LysAspAsnArgThrSerTyrGluCysAlaGluHisSerAlaProIleLeu 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1180 CGTGGGGCCACGTGGTGGTGCAGAAACCCAG..... 1148
91  sValLeuHisAsnGlnThrGlyLysTyrArgTyrAspAlaGluThrG 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1147 .....TCAACTGGGATCCAGAGACTG 1126
108  InGlyTrpIleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIle 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1125 TCGGCTCATACATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1076
125  ProGlyGlyTyrValAlaSerArgSerGlyGlyLysLeuLeuGlyPhe 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1075 CCTGGAGGATTTATCTGCCAANAATTCAGCAGCAACAGGGCTTTGGCTT 1026
141  sGlyIlePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAla 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1025 TGCCATTGTGGCTACCTCCACCCTAAACATGTTGATCCCTTCAGCAGGCC 976
158  sPheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGly 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
975 GCGTTCACATATGGCTGTGTCTATCTTCGTGAGGATCCCTTCAGGATTGGT 926
175  GluGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaPr 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
925 GAGGGGGTCACATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
191  oProLeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGln 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
875 TCCCTTAGACGGAGTCGCTGGCAACACAGCAGCTTTTGGGTTCTCTATG 826
208  euGlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMet 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825 CTGGGGCGGTGGTGGCATGCGCTTGGCTGGGTCCTTGTGAGTATTCA 776
225  AsnTrpThrTyrValPheTyrPhePheGlyIleValGlyIleIleTrpPh 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
775 GGATGGAGTTCTGCTCTCTATGCTATGGCAGCTTCGGGATCTTTTGGA 726
241  eIleLeuTrpIleCysLeuValSerAspThrProGluThrHisLysThr 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
725 CCGTGTCTGGTGTGCTGTCTCTATGATCACCAGCAGCTGCACCCAGCA 676
258  leThrProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnLeu 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
675 TCTCTGAGGAGGAGCGCAATACATTGAGGATGCCATCGGGGAGAGCGCC 626
275  SerSerGlnLysSerVal.....ProTrpIleProMetLe 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
625 AAGCTCATGACCCCTGTTACCAAGTTTAAACACACCCCTGGAGCGCTTCT 576
286  uLysSerLeuProLeuTrpAlaIleValValAlaHisPheSerTyrAsn 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
575 TACGTCCATGCCCTCTATGTCATCATCTGTCGGAACCTTTTCCGCGAGCT 526
303  rpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGluVal 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 GGACCTTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 476
320  LeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLe 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
475 TTCGGCTTGAGATCAGCAAGGTGGGCTGTGTGCGGCGCTGCCTCAGCT 426
336  uGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeu 353
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 TGTCTATGACCATCATCGTACCATTGGAGGCCAGATCGCTGACTTTTTCG 376
353  rgAlaArgTrpAsnPheSerThrLeuTrpValArgValPheSerLeu 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 CGAGTCGTGCATATGTCTTCCATCTACCAACGTGCGAAGACCTCATGACGTC 326

```

```

370  IleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGl 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325 GGGGTTTTGGGATGGGAAGCCACGCTGCTGTGGTGGTGGATAC...TC 279
386  yCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuG 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 GCATCCCAAGGGCGTGGCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 229
403  lyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaPro 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
228 GTGGCTTGGCCATCTCTGGTTTACGTGAACCACTTGCACATCGCCCT 179
420  SerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIlePr 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178 CGCTATGCCAGCATCTTGATGGCATTTCCCAATGGCGTGGGCACACTGTC 129
436  oGlyMetIleGlyProIleIleAlaArgSerLeuThrProGluAsnThrI 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 TGGGATGGTGGCCCCATCATCGTGGTGCATGACCAAGCACAAGACGC 79
453  leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 GGGAGGAGTGGCAGTACGTGTCTCTCATAGCTCCCTGGTGACACTACGC 29
470  GlyAlaIlePheThrLeuPheAla 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28 GGTGCATCTCTTATGGGCTCTTTGCT 2

```

seq_name: gb_est2:BI860521

seq_documentation_block:

LOCUS BI860521 736 bp mRNA linear EST 10-OCT-2001
DEFINITION 603386787F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5',
mRNA sequence.
ACCESSION BI860521
VERSION BI860521.1 GI:16001268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12008 row: m column: 21
High quality sequence stop: 713.

FEATURES
source

Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5395772"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: Sali; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 216 a 142 c 147 g 231 t
ORIGIN

alignment_scores:
 Quality: 879.50 Length: 224
 Ratio: 4.557 Gaps: 2
 Percent Similarity: 86.161 Percent Identity: 75.000

alignment_block:
 US-09-776-865-4 x BF539146 ..

Align seg 1/1 to: BI860521 from: 1 to: 736

79 ThrSerTyrGluCysAlaGluHisSerAlaProIleLysValLeuHisAs 95
 1 ACTTCCAAGGGCTGTCAGAGCATCTGCTCCCATAAAGTTTCATCATAA 50
 95 nGlnThrGlyLysLysTyrArgTrpAspAlaGluThrGlnGlyTrpIleL 112
 51 TCAACAGGGTAAGAAGTACCAATGGGATGCGAGAACTCAAGGATGGATTC 100
 112 euGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyr 128
 101 TCGGTCTCTTTTATGGGTACATCATCACAGATTCCTGGAGGATAT 150
 129 ValAlaSerArgSerGlyGlyLysLeuLeuGlyPheGlyIlePheAl 145
 151 GTTCCAGCAAAATAGGGGGAATGCTCTAGGATTTGGGATCTCTGG 200
 145 aThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPheGlyValG 162
 201 CACTGCTCTCTCCACCCTGTTCACTCCCATGCTGCGAGATTTAGGAGTTG 250
 162 lYAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluGlyValThr 178
 251 GACCACCTATTGTACTCAGAGCACTAGAGGACTAGGAGGAGGTTGTACA 300
 179 TyrProAlaMetHisAlaMetTrpSerSerTrpAlaProLeuGluArg 195
 301 TTTCCAGCATGATGCGATGCTGCTCTGCGCTCCCTCTTGAAG 350
 195 gSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValV 212
 351 AAGCAAACTCTTAGCATTTTCATATGAGGAGGACAGCTTGGGACAGTAA 400
 212 alSerLeuProLeuSerGlyValIleCysTyrTyrMetAsnTrpThrTyr 228
 401 TTTCTCTCTCTCTCTGGAATATTTGCTACTATATGAATTTGACATTAT 450
 229 ValPheTyrPhePheGlyIleValGlyIleIleTrpPheIleLeuTrpI 245
 451 GTCITCTACTTTTGGTACTATTGGAATATTTGTTCTTTTGGGAT 499
 245 eCysLeuValSerAspThrProGluThrHisLysThrIleThrProTyrG 262
 500 CTGGTTAGTAGTACACACACACACACACACACACAGAAATTTCCCATATG 549
 262 luLysGluTyrIleLeuSerSerLeuLysAsnGlnLeuSerSerGlnLys 278
 550 AAAAGGAATACATCTTTTCATCATTAAGAAATCAG.....584
 279 SerValProTrpIleProMetLeuLysSerLeuProLeuTrpAlaIleVa 295
 585GTATGGACTTTG...596
 295 lValAlaHisPheSerTyrAsn 302
 597ACACATTTTAAATAAT 614

seq_name: gb_est2:BF539146

seq_documentation_block:
 LOCUS BF539146 895 bp mRNA linear EST 11-DEC-2000
 DEFINITION 602053009F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4192035 5',
 mRNA sequence.

ACCESSION BF539146
 VERSION BF539146.1 GI:11626527
 KEYWORDS EST.
 SOURCE house mouse
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9522 row: b column: 04
 High quality sequence stop: 710.

BASE COUNT 198 a 232 c 263 g 202 t

ORIGIN

alignment_scores:
 Quality: 868.00 Length: 280
 Ratio: 3.678 Gaps: 8
 Percent Similarity: 84.286 Percent Identity: 70.714

alignment_block:
 US-09-776-865-4 x BF539146 ..

Align seg 1/1 to: BF539146 from: 1 to: 895

9 AlaProSerAspGlyGluGlySerAspArgThrProLeuLeuGlnArg 25
 58 GCGGAAACGACGATGAGGAGCTCGGACAGCACCCTCTCTCGCGGG 107
 25 gAlaProArgAlaGluProAlaProValCysSerAlaArgTyrAsnL 42
 108 CGCCGCGACAGCCGAGCGGCTCCAGTGTCTGCTCTGCTCGGTACACT 157
 42 euAlaPheLeuSerPhePheGlyPheValLeuTyrSerLeuArgVal 58
 158 TAGCGAT.TTGGGCTTCTGTGGTTCTCTGTTCTCTATGCTTACGGGTG 206
 59 AsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThrAlaLys 75
 207 AACCTGAGTCT.GCGTTAGTGGACATGGTAGATTCAATACAACTCTGAC 255
 75 sAspAsnArgThrSerTyrGluCysAlaGluHisSerAlaProIleLysV 92
 256 TGATAATAGAACGCTCTAAGGAGTGTGCGGAACATTTCTGCCCATAAAG 305
 92 alLeuHisAsnGlnThrGlyLysTyrArgTrpAspAlaGluThrGln 108
 306 TTCACCAATACACAGGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 355
 109 GlyTrpIleLeuGlySerPhePhe.TyrGlyTyrIleIleThrGlnIleP 125

```
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 405
125 roGlyGlyTyrValAsnSerGlyGlyLysLeuLeuGlyPhe 141
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 455
406 CGGTGGGTGATCTCCAGAGGTCGGAGGAAGCTCTCTGGGCGTG 455
142 GlyIlePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAs 158
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 505
456 GGCACTTTAGGCACCTCGCTCTTACCCCTGTTACACCGCTGGCGCAGA 505
158 pPheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGlyGly 175
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 554
506 CTTAGGCGTGTGACTCTCTCGTGTG.CTTAGAGCGCTGGAAGACTGGGAG 554
175 LuGlyValThrTyrProAlaMetHisAlaMetTyrSerTyrAlaPr 191
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 604
555 AGGTGTTTACGTTTCCAGCTATGACCGCATGTGCTCTCTGGGCTCC 604
191 oProLeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnL 208
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 654
605 CCCCTGGAAGAAGCAAGCTTCTTACCATTTCTATGCGGCGGACAGC 654
208 euGlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMet 224
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 704
655 TTGGACAGTATCTACTCTCTCTGCGGAATAATATGCTACTATATG 704
225 AsnTrp.ThrTyrValPhe.TyrPhePheGlyIleValGlyIleLeuTrp 240
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 751
705 AACTGGGACTTAAGTCTTCTAAC...TTGGACTAGCGGAGTGGCGGT 751
241 PheLeuLeuTrpIleCysLeuValSerAspThrProGluThrHisLysTh 257
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 796
752 ...ACACAATGGGAGGCGGAGCGGTGATACCCCGAAATTCGAGAA... 796
257 rIleThrProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnL 274
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 836
797 ....TTCCCTTTTGGAGACAC.....ACTGGTGCCTCTAAACAAACGTT 836
274 euSerSerGlnLysSerValProTyrIlePro 284
|||||GGGTGGATTCTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 868
837 TGGAAAGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 868
```

seq_name: gb_estl:BB537525

seq_documentation_block: 645 bp mRNA linear EST 26-OCT-2001
LOCUS BB537525 RIKEN full-length enriched, 0 day neonate eyeball Mus
DEFINITION musculus cDNA clone E130006F23 3' similar to AJ387747 Homo sapiens
mRNA for sialin, mRNA sequence.

ACCESSION BB537525 GI:16446752

VERSION BB537525.2

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 645)

Akakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
, M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Jul 31, 2000 this sequence version replaced gi:9593025.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,K., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

FEATURES

Location/Qualifiers

1..645

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone_lib="E130006F23"

/clone_lib="RIKEN full-length enriched, 0 day neonate

eyeball"

/tissue_type="eyeball"

/dev_stage="0 day neonate"

/lab_host="DH10B"

/note="Site_1: Sall; Site_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'

GAGAGAGAGCGCGCGCAACTCGAGTTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGAGATTCTCGAGTTAATAATATCCCGCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified

pBluescript KS(+) after bulk excision from Lambda FLC 1."

BASE COUNT 153 a 154 c 160 g 177 t

ORIGIN

alignment_scores:

Quality: 839.50 Length: 191

Ratio: 4.716 Gaps: 2

Percent Similarity: 93.194 Percent Identity: 83.770

alignment_block:

US-09-776-865-4 x BB537525

Align seg 1/1 to: BB537525 from: 1 to: 645

49 GlyPhePheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuVa 65

```
|||||
20 GGGTCTCTGCTCTATGCTTANCGGTGAACCTGAGTGTGCGTTAGT 69
65 lAspMetValAspSerAsnThrTrpAlaLysAspAsnArgThrSerTyrg 82
70 GGACATGGTAGATTCAATACAACTCTGACTGATATAGAACGCTAAGG 119
82 luCysAlaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGly 98
120 AGTGTGGGAACATCTGCCCCATAAAAGTTCACCAACAATCACACAGT 169
99 LysLysTyrgArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPh 115
170 AAAAGTACAAAGTGGGATGAGAACTCAAGGGTGGATTCTCGGCTCTT 219
115 ePheTyrgLysTyrlleIleThrGlnIleProGlyGlyTyrgValAlaSerA 132
220 TTTTACGGCTACATCTGCCAGATCCCGGTGGGTGATATGCCAGCA 269
132 rgSerGlyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIle 148
270 GGGTCGGAGGAAAGCTGCTGGGCTGGGCATCTTAGGCACCTCCGTC 319
149 PheThrLeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuVa 165
320 TTCACCTGTTCACACCGCTGGCGCAGACTTAGGCGGTGGTACTCTCGT 369
165 lAlaLeuArgAlaLeuGluGlyLeuGlyGlu.....Gly. 176
370 TGTGCTTAGAGCGCTGGAAGAGCTGGGAGAGAAATATCCACCCAGGT 419
177 ValThrTyrgProAlaMetHisAlaMetTrpSerSerTrpAlaProLe 193
420 GTTACGTTTTCAGTATGACGCCATGTGTCTTCTTGGGCTCCCTCT 469
193 uGluArgSerLysLeuSerIleSerTyrgAlaGlyAlaGlnLeuGlyT 210
470 GGAAGAAGCAAGCTTCTTACCATTTCCTATGCGGGAGCACAGCTGGGA 519
210 hrValValSerLeuProLeuSerGlyValIleCysTyrgTyrgMetAsnTrp 226
520 CAGTGATCTCACTCTCTTCCGGAATATATGCTACTATATGAAGTGG 569
227 ThrTyrgValPheTyrgPhePhe 233
570 ACTTACGCTCTCTATCTTTT 590
```

seq_name: gb_est2:BF124137

```
seq_documentation_block:
LOCUS      BF124137
DEFINITION 601760838F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4024064 5',
            mRNA sequence.
ACCESSION  BF124137
VERSION     BF124137.1 GI:10963177
KEYWORDS   EST.
SOURCE      house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 904)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
Plate: LLAM9283 row: k column: 09
High quality sequence stop: 680.
FEATURES
    source
        1..904
            /organism="Mus musculus"
            /strain="C57/B6"
            /db_xref="taxon:10090"
            /clone="IMAGE:4024064"
            /clone_lib="NCI_CGAP_Mam5"
            /tissue_type="tumor, gross tissue"
            /dev_stage="7 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
            Library constructed by Life Technologies. Investigators
            providing samples: Lothar Hennighausen/Robin Humphreys,
            NIH"
BASE COUNT  193 a 214 c 276 g 221 t
ORIGIN
alignment_scores:
    Quality: 837.00      Length: 223
    Ratio: 4.024        Gaps: 0
    Percent Similarity: 93.274    Percent Identity: 82.063
alignment_block:
    US-09-776-865-4 x BF124137 ..
Align seg 1/1 to: BF124137 from: 1 to: 904
9 AlaProSerAspGlyGluGlySerAspArgThrProLeuLeuGlnAr 25
54 GCGGAAACGACGATGAGGAGCTCGACAGCACCCGCTCTGCCGGG 103
25 gAlaProArgAlaGluProAlaProValCysSerAlaArgTyrrAsnL 42
104 CGCCCGCAGACCCGAGCGCTCCAGTGTGCTGCTGCTGCTGCTACT 153
42 euAlaPheLeuSerPhePheGlyPhePheValLeuTyrgSerLeuArgVal 58
154 TAGCGAT..TTGGCGTCTCTGTGGTTCTTCGTCTCTATGCTCTACGGGTG 202
59 AsnLeuSerValAlaLeuValAspMetValAspSerAsnThrTrpAlaLy 75
203 AACCTGAGTGT..GGTTAGTGGACATGTTAGATTCAATACAACTCTGAC 251
75 sAspAsnArgThrSerTyrgLysAlaGluHisSerAlaProIleLysV 92
252 TGATAATAGAACGCTCTAAGGAGTGTGCGGAACATCTGCCCCCAATAAAG 301
92 allLeuHisAsnGlnThrGlyLysTyrgArgTrpAspAlaGluThrGln 108
302 TTCACCACAATCACAGGTAAAGATACAGTGGGATGAGAACTCAA 351
109 GlyTrpIleLeuGlySerPheTyrgTyrlleIleThrGlnIlePr 125
352 GGGTGGATTCTCGGCTCTTTTTCACGGCTACATCGTACCCAGATTCC 401
125 oGlyGlyTyrgValAlaSerArgSerGlyGlyLysLeuLeuGlyPheG 142
402 CGGTGGGTACATTGCCAGCAGGCTCGGAGGAGAGCTGCTGCTGGGCTGG 451
142 llylePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAsp 158
452 GCATCTTAGGACCTCTTACCTCTTACCTGTTCACCCGCTGGCGCGCAGAC 501
159 PheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGlyLeuGlyGln 175
502 TTAGGCGGTGACTCTCGTGTGCTTTAGAGCGCTGGAAGGAGCTGGGAGA 551
175 uGlyValThrTyrgProAlaMetHisAlaMetTrpSerSerTrpAlaProp 192
|||||
```


142 lyllePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAsp 158
 486 GCATCTTAGGCACCTCCGCTTCACCTGTTCACACCGCTGGCGCAGAC 535
 159 PheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyG 175
 536 TTAGGCGTGGTACTCTCGTGTGCTTAGAGCGCTGGAAGGACTGGGAGA 585
 175 uGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaA 191
 586 GGGTGTAGCTTCCAGCTATGCACGCCATGCTGCTTCCCTGGCTCCCC 635
 192 ProLeuGluArg.SerLysLeuLeuSerLysSerTyrAlaGlyAla 206
 636 CCTCTGGAAGAAGCGAAGCTTCTTACCATTTCCCTATGCGGGGAGC 681

seq_name: gb_est2:BG541099

seq_documentation_block:
 LOCUS BG541099 893 bp mRNA linear EST 03-APR-2001
 DEFINITION 602570265F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4694680 5',
 mRNA sequence.

ACCESSION BG541099
 VERSION BG541099.1 GI:13533332
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 893)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1CM1519 row: a column: 17
 High quality sequence stop: 549.
 Location/Qualifiers

FEATURES
 source
 1..893
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:4694680"
 /clone_lib="NIH_MGC_77"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgcc); Site_2: SfiI (ggccattatgcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGCGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGCGGCGCAGATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 BASE COUNT 236 a 169 c 214 g 274 t
 ORIGIN

alignment_scores:
 Quality: 808.00 Length: 324
 Ratio: 3.848 Gaps: 8
 Percent Similarity: 64.815 Percent Identity: 54.321

alignment_block:
 US-09-776-865-4 x BG541099 ..

Align seg 1/1 to: BG541099 from: 1 to: 893
 162 GlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluGlyValPh 178
 1 GGACCACTCATTTGTACTCAGACACTAGAGGACTAGGAGAGGGTGTAC 50
 178 rTyrProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGluA 195
 51 ATTTCAGCCATGCATGCCATGTGGTCTTCTTGGGCTCCCTCTTGA 100
 195 rGSerLysLeuLeuSerLysSerTyrAlaGlyAlaGlnLeuGlyThrVal 211
 101 GAAGCAAACTTCTTAGCATTTTCATATGCAGGACACAGCTGGGACAGTA 150
 212 ValSerLeuProLeuSerGlyValIleCysTyrTyrMetAsnTrpThrTy 228
 151 ATTTCCTCTCTCTTCTGGAATAATTTGCTACTATATGAATGGACTTA 200
 228 rValPheTyrPhePheGlyIleValGlyIleLeIleTrpPheIleLeuTri 245
 201 TGTCTTCTACTTTTGGTACTATTGGAATATTG...GTTTCTTTGTGGA 247
 245 leCysLeuValSerAspThrProGluThrHisLysThrIleThrProTyr 261
 248 TCTGGTTAGTTAGTACACACACACAAAAACACAGAGAATTTCCCATTA 297
 262 GluLysGluTyrIleLeuSerSerLeuLysAsnGlnLeuSerSerGlnLy 278
 298 GAAAGGATACATCTTTCATCATTAAGAAATCAG..... 333
 278 sSerValProTriPleProMetLeuLysSerLeuProLeuTriPAlaIleV 295
 333 333
 295 alValAlaHisPheSerTyrAsnTrpThrPheTyrThrLeuLeuThrLeu 311
 333 333
 312 LeuProThrTyrMetLysGluValLeuArgPheAsnIleGlnGluAsnGl 328
 334AATGG 338
 328 yPheLeuSerAlaValProTyrLeuGlyCysTrpLeuCysMetIleLeuS 345
 339 GTTTTATCTTCATTCCTTATTAGGCTCTTGGTTATGATGATGATCTCT 388
 345 erGly.GlnAlaAlaAspAsnLeuArgAlaArgTrpAsnPheSerThrle 361
 389 CTGGCCCAAGCTGCTGACAAATTTAAGGCAAAATGGAATTTTCACTTT 438
 361 utrPValArgArgValPheSerLeuIleGlyMetIleGlyProAlaIleP 378
 439 ATGCTGTCGCAAAATTTTAGCTTATAGGAATGATGGACCTGCAGTAT 488
 378 heLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaValAla 394
 489 TCCTGGTAGCTGCTGGTTCATTTGGTGTGATTATTTCTTGGCCGCTGT 538
 395 PheLeuThrIleSerThrThrLeuGlyPheCysSerSerGlyPheS 411
 539 TCTAATACTATATCAACAACACT...GGAGGCTTTTGGTCTCTCTGGATTAG 585
 411 rIleAsnHisLeu.AspIleAlaProSerTyrAlaGlyIleLeuLeuGly 427
 586 ATCCACCCATTTGGGATATG..... 605
 428 IleThrAsnThrPheAlaThrIleProGlyMetIleGlyProIleIleAl 444
 606GTGCTTCCACACACTGTTGGGAGAAATGGGAAC 637
 444 aArgSerLeuThrProGluAsnThrIleGlyGluTrpGlnThrValPheC 461
 638 CGGGTCTCTT.....ATGGTGGGGGAATTAAGTCT..... 668

/organism="Canis familiaris"
 /db_xref="taxon:9615"
 /clone_lib="Canine Brain cDNA Library"
 /sex="Male"
 /note="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site: 1:
 NotI; Site: 2: SalI; Tissue was taken from the frontal,
 occipital, temporal and parietal lobes, olfactory bulb,
 hippocampus, cerebellum, thalamus, hypothalamus, midbrain
 , pons, and medulla."
 BASE COUNT 164 a 123 c 127 g 211 t 8 others
 ORIGIN

alignment_scores:
 Quality: 796.00 Length: 222
 Ratio: 4.303 Gaps: 5
 Percent Similarity: 83.333 Percent Identity: 72.072

alignment_block:

US-09-776-865-4 x BI817031

Align seg 1/1 to: BI817031 from: 1 to: 633

237 GlyllelelrrpPheilleleuTrpilleCysleuValSerAspThrProgl 253
 |||||
 1 GGCAATCATTTGGTTATTTATTTATGGATCTTTTATGTTAGTGAACCTCCAGA 50
 253 uThrHisLysThrIleThrProTyrgluLysgluTyrrilleleuSerSerL 270
 |||||
 51 GACTCACAGAACATCTCCCGGAGGAAGAATATATCTTTTCATCAT 100
 270 euLysAsnGlnLeuSerSerGlnLysSerValProTrrpilleProMetLeu 286
 |||||
 101 TAAAAAATCAGCTTTCATCACAGAAGTCAGTGCATGATGCCATGCTA 150
 287 LysSerLeuProLeuTrpAlaIleValAlaIleHisPheSerTyrrAsnTr 303
 |||||
 151 AAATCACTGCGCGTTTGGGCTATTGTAGTAGCATATTTTCTTCAACTG 200
 303 pThrPheTyrrThrLeuLeuThrLeuProThrTyrrMetLysGluValL 320
 |||||
 201 GACTTTTACACTTTATGACGTTATGCTTACATGAGGAGATCC 250
 320 euArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrrLeu 336
 |||||
 251 TAAGGTTCAATGTTCAAGAGAATGGCTTCTATCTGCACCTGCCTACTT 300
 337 GlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuAr 353
 |||||
 301 GGCTGTTGGTTATGATGATCTTCTGGTCAAGCTGCTGACAAATTTAAG 350
 353 gAlaArgTrpAsnPheSerThrLeuTrpValArgValPheSerLeuI 370
 |||||
 351 GGCAAAATGGNAATTTTCAACAATATGTGCCGAGAGATTTTATGCCCTAA 400
 370 leGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGly 386
 |||||
 401 TAGGAATGATTGGACCTGGGTATTCCTGGTAGCGCTGGATTATAGGT 450
 387 CysAspTyrrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuG 403
 |||||
 451 TGTGACTATTATTCATGGCGGTGGGT...CTTAACCATATCACACACTGG 497
 403 lYGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaPro 419
 |||||
 498 GAGCTTTTCTCTTNT..... 514
 420 SerTyrrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleP 436
 |||||
 515GGATTAAACATCAACATCTGGATGATGTC 543
 436 roGlyMetIleGlyProIleIleAlaArgSerLeuThrProGluAsnThr 452
 |||||

544 CTTTCATGCTGGATCTCTG.....GGCATCACAAATCATTTGCACT 584
 453 lleGlyGluTrp 456
 ::: |||||
 585 TTTTCAGGAATGG 596

seqname: gb_est2.BE867611

seqdocumentation_block:

IRUS BE867611 576 bp mRNA linear EST 20-OCT-2000
 DEFINITION 601443127F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847279 5',
 mRNA sequence.

ACCESSION BE867611

VERSION BE867611.1 GI:10316387

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 676)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: L1AM9561 row: e column: 08

High quality sequence stop: 676.

FEATURES

source

1. 676

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3847279"

/clone_lib="NIH_MGC_65"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 1.8 kb. Library constructed by Life

Technologies."

BASE COUNT 145 a 171 c 184 g 176 t

ORIGIN

alignment_scores:

Quality: 782.00 Length: 192

Ratio: 4.418 Gaps: 2

Percent Similarity: 92.188 Percent Identity: 79.167

alignment_block:

US-09-776-865-4 x BE867611

Align seg 1/1 to: BE867611 from: 1 to: 676

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlyGly 17
 |||||
 103 ATGAGGTCCTCGGTTTCGAGACCTGGCCGGAACGATGCGGAGGAGAC 152
 17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaPro 34
 |||||
 153 GGACCGCAGCGCTCTTCTACCGGGCGGCCACGGCGCGAGCGCTCCAG 202
 34 alCysCysSerAlaArgTyrrAsnLeuAlaPheLeuSerPheGlyPhe 50
 |||||
 203 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 252
 51 PheValLeuTyrrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67

us-09-776-865-4.p2n.rst

183 isAlaMetTrpSerSerTrpAla 190
 653 TTGCCCATGTGGTCTTCTTGGGCT 675